Draft version 1.0, 27/05/2019. This paper has not been peer reviewed.

# Memory suppression and its deficiency in psychological disorders: A focused meta-analysis

Davide F. Stramaccia[a], Katharina M. Rischer[b], Jonathan M. Fawcett[c], and Roland G. Benoit[a]

[a]Max Planck Research Group - Adaptive Memory, Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany.
[b]Faculty of Language and Literature, Humanities, Arts and Education, Université du Luxembourg Maison des Sciences Humaines 11, Porte des Sciences L-4366 Esch-sur-Alzette, Luxembourg.
[c]Department of Psychology, Memorial University of Newfoundland, 230 Elizabeth Ave, St. John's, NL A1C 5S7, Canada.

**Word count: 9043**

Contacts
Davide F. Stramaccia: stramaccia@cbs.mpg.de
Roland G. Benoit: rbenoit@cbs.mpg.de

## Abstract

It is hotly debated whether suppressing the retrieval of unwanted memories constitutes a beneficial mechanism that causes forgetting. Here, we scrutinize the evidence for such suppression-induced forgetting (SIF) and examine whether it is deficient in psychological disorders characterized by intrusive thoughts. Specifically, we performed a focused meta-analysis of studies that have used the *Think/No-Think* procedure to test SIF in individuals either affected by psychological disorders or exhibiting high scores on related traits. First, our analysis of the control samples ($N$ = 534) indicated that avoiding retrieval indeed leads to reliable forgetting in healthy participants. Overall, the effect size was moderate to small (*SMCC* = 0.31, 95% CI [0.16, 0.45]) and remained significant after attempting to account for publication bias. However, moderator analyses revealed that this effect varied according to the exact mechanism that participants were instructed to engage, with the greatest effect size observed for direct retrieval suppression (*SMCC* = 0.63, 95% CI [0.36, 0.90]). Second, we found no evidence for SIF in the clinical/sub-clinical samples ($N$ = 534, *SMCC* = 0.07, 95% CI [-0.13, 0.28]). Critically, SIF in these samples was significantly smaller than in the respective control samples (*SMD* = 0.26 (95% CI [0.06, 0.47]). This deficiency was particularly pronounced when participants were instructed to apply direct retrieval suppression mechanism. These results suggest that intact suppression-induced forgetting is a hallmark of psychological well-being, and that inducing more specific suppression mechanisms fosters voluntary forgetting.


**Keywords:** suppression; involuntary retrieval; cognitive control; anxiety; depression.

# 1. Introduction

*In the practical use of our intellect, forgetting is as important a function as remembering.*

William James, 1892

Forgetting is often regarded as a deficiency of our memory systems, where attempts to retain or retrieve information are met with failure. However, under many circumstances forgetting can be characterized as an adaptive force that shapes our memory, for instance by updating or discarding information that has become irrelevant – or even outright unwanted (Nørby, 2015). Accumulating evidence suggests that such forgetting can be under intentional control: concerted attempts at preventing cued memories from entering awareness can subsequently make it more difficult to voluntarily retrieve these suppressed memories and eventually cause forgetting (Anderson & Hanslmayr, 2014). In essence, such suppression-induced forgetting (SIF) (Anderson & Huddleston, 2014; Hertel & McDaniel, 2010) may serve the purpose of preventing our minds from being at the mercy of involuntary retrieval.

We here conducted a focused meta-analysis to scrutinize whether it is possible to foster forgetting *intentionally*. We were particularly interested in gauging whether such intentional forgetting may be a hallmark of psychological well-being (see also Engen and Anderson, 2018) and thus be deficient in people suffering from disorders characterized by intrusive thought.

A deficiency in controlling one's memories and thoughts may be at the heart of several psychological disorders (e.g., Goschke, 2014; Hertel, 1997, 1998, 2007; McTeague, Goodkind, & Etkin, 2017). Perhaps most prominently, post-traumatic stress disorder (PTSD) is characterized by intrusive memories and unintentional re-experiencing (Brewin, 2014; Ehlers, Hackmann, & Michael, 2004; Hackmann, Ehle, Speckens, & Clark, 2004). Indeed, this feature of PTSD has been recognized as one of its defining aspects in both the Diagnostic and Statistical Manual of Mental Disorders (5th ed.; DSM–5; American Psychiatric Association, 2013) and the International statistical classification of diseases and related health problems (11th ed.; ICD; World Health Organization, 2018).

The intrusiveness of memories in PTSD may result from an impaired ability to keep unwanted memories at bay (Ehlers et al., 2004; Hackmann et al., 2004). Patients suffering from this condition tend to seek help after intrusive memories would already had time to become strongly consolidated, thus highlighting the importance of understanding the retrieval processes that support the intrusions (Marks, Franklin, & Zoellner, 2018). Similarly, intrusive negative thoughts also constitute central symptoms of affective disorders such as anxiety (Kircanski, Johnson, Mateen, Bjork, & Gotlib, 2016) and depression (Kircanski, Joormann, & Gotlib, 2012). These intrusive thoughts have also been suggested to arise from the involuntary retrieval of previously experienced or imagined episodes (Iyadurai et al., 2018; Visser, Lau-Zhu, Henson, & Holmes, 2018).

If involuntary retrieval constitutes a core symptom of several psychological disorders (e.g., Goschke, 2014; Hertel, 1997, 2007; McTeague et al., 2017), then the ability to control such unwanted memories may constitute a mechanism that promotes well-being (Benoit, Davies, & Anderson, 2016; Depue, Curran, & Banich, 2007; Engen and Anderson, 2018; Joormann, Hertel, LeMoult, & Gotlib, 2009; Visser et al., 2018). We here test this account by reporting two meta-analyses on voluntary memory suppression. The first tries to establish whether it is possible to intentionally forget unwanted memories by controlling their retrieval. The other analysis examines the hypothesis that an impairment in this ability constitutes a vulnerability towards developing clinical disorders. In this case, we expect that such intentional forgetting may be deficient in people suffering from psychological disorders that are characterized by intrusive thought.

The behavioral procedure typically used to elicit SIF has been called *Think/No-Think* (Anderson & Green, 2001). In this procedure (Figure 1), participants first learn to associate pairs of cues and targets (e.g., *TOMATO – VEST*), so that they can retrieve the target (*VEST*) upon presentation of its cue (e.g., *TOMATO*). Participants then enter the critical *Think/No-Think* phase, where they are shown a subset of the cues. For some of these cues, participants have to covertly rehearse the associated target (i.e., *recall* items). For other cues, participants need to actively prevent the associated target from coming to mind (i.e., *suppress* items). Each of those cues are presented several times, so to provide multiple opportunities for memory-control mechanisms to be deployed. A

number of cues are not shown at all during this phase (i.e., *baseline* items), and serve to assess baseline memory performance in a following test phase. On that test, participants are instructed to recall each response (e.g., *VEST*) upon presentation of its specific cue (e.g., *TOMATO*), irrespective of previous instructions. Typically, participants are impaired at retrieving previously suppressed memories as indicated by worse memory accuracy for suppress than for baseline items. This finding of below-baseline memory accuracy is considered an index of SIF.

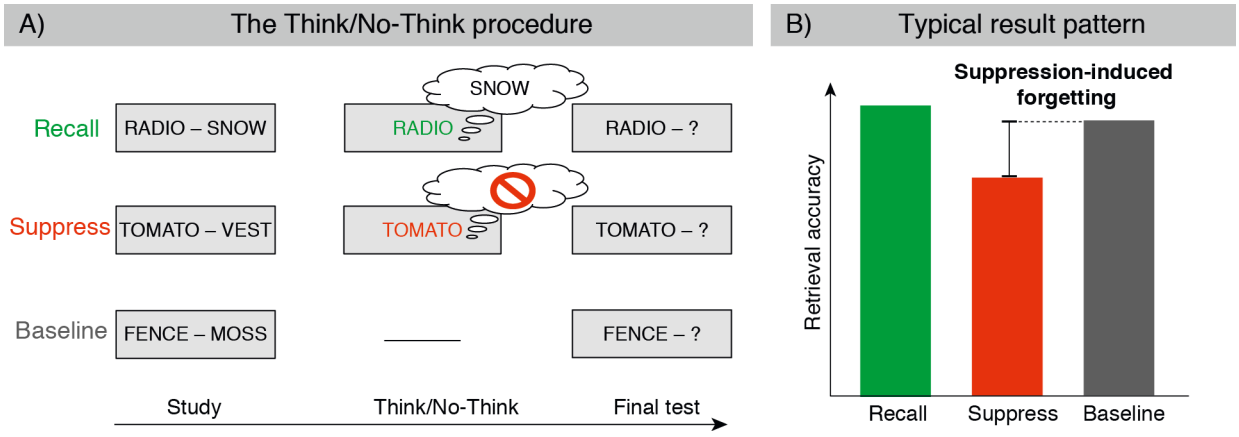


**Figure 1.** Panel A) Overview of the *Think/No-Think* procedure. In the initial study phase, participants encode associations of cues (e.g., RADIO) and targets (e.g., SNOW). They then enter the critical *Think/No-Think* phase, in which they repeatedly encounter most of the cues. For some of the cues (here for those presented in green), participants attempt to recall the associated targets (Recall items). For other cues (here for those presented in presented in red), their task is to prevent the associated target memory from coming to mind (Suppress items). A third of the targets that they had also initially learned are not cued during this phase (Baseline items). On a final test, participants are asked to remember all targets given their respective cues, irrespective of the previous instructions. Panel B) Typical retrieval accuracy on the final test. Participants are generally better or similarly capable at remembering previously rehearsed Recall than Baseline targets. Critically, participants are typically worse at retrieving previously Suppressed targets than Baseline targets. We refer to this latter finding as suppression-induced forgetting.

Though there has been accumulating evidence for SIF over the last 15 years, this effect has not universally been replicated (e.g., Algarabel, Luciano, & Martínez, 2006; Bergström, Velmans, de Fockert, & Richardson-Klavehn, 2007); Bulevich, Roediger, Balota, & Butler, 2004; Mecklinger, Parra, & Waldhauser, 2009; Wessel, Wetzels, Jelicic, & Merckelbach, 2005). A major goal of this analysis is thus to determine the reliability

and effect size of SIF. This is particularly important, as without confidence in the reliability or magnitude of SIF, it would be difficult to evaluate related deficits in clinical populations. Prior to evaluating the relation between the SIF effect and mental health, we will thus estimate SIF in non-clinical individuals.

Some of the inconsistencies in the literature may reflect important study differences with respect to the exact mechanisms that people engaged to prevent unwanted retrieval. While initial studies were somewhat agnostic regarding the employed processes (e.g., Anderson & Green, 2001), there is now evidence for two specific suppression mechanisms. On one hand, people can prevent recall by stopping the retrieval process altogether (Benoit & Anderson, 2012; Bergström, de Fockert, & Richardson-Klavehn, 2009; Gagnepain, Henson, & Anderson, 2014). This mechanism, *direct retrieval suppression*, has been associated with an inhibitory top-down modulation of the hippocampus that originates from the right dorsolateral prefrontal cortex (Benoit & Anderson, 2012; Gagnepain et al., 2014). The other mechanism, *thought substitution*, requires participants to retrieve an alternative memory when faced with a cue to an unwanted memory. This substitute memory then occupies the limited focus of awareness and thus prevents the unwanted memory from coming to mind (Benoit & Anderson, 2012; Bergström et al., 2009; Hertel & Calcaterra, 2005). Thought substitution has been associated with memory selection processes supported by the left ventrolateral prefrontal cortex (Benoit & Anderson, 2012). Critically, both of these mechanisms have been shown to cause forgetting (Benoit & Anderson, 2012; Bergström et al., 2009; Hertel & Calcaterra, 2005), though there is evidence that they may not always be equally efficient (Bergström et al., 2009). We will thus examine whether SIF in healthy individuals varies according to the induced suppression mechanism.

Turning to clinical populations, there is indeed evidence for impaired SIF, for example in PTSD (Sullivan et al., 2019; Waldhauser et al., 2018; Catarino, Küpper, Werner-Seidler, Dalgleish, & Anderson, 2015). However, the reliability of such a deficiency in clinical populations is still uncertain, because several studies did not directly observe impaired SIF (as compared with SIF in the respective healthy control group). Instead, these studies inferred memory control impairments from other between-groups differences that are less stringent indices of intentional forgetting. These include higher

recall of *suppress* items (e.g., Hertel & Gerstle, 2003), impaired recall of *baseline* items (e.g., Hertel & Mahan, 2008; Joormann, Hertel, Brozovich, Gotlib, 2005), and different patterns of neural activation during the *Think/No-think* phase as revealed by functional MRI (Sacchet et al., 2017).

To shed light on these issues, we meta-analyzed studies that compared clinical and sub-clinical samples with healthy controls on SIF as elicited by the *Think/No-Think* procedure. As detailed above, we predicted a significant SIF effect for healthy control groups, as well as a significant difference between healthy control groups and clinical and sub-clinical samples.

Furthermore, we explored the effects of a few important features that might influence the magnitude of SIF. First, we assessed the impact of providing different instructions that are either targeted at inducing specified mechanisms (i.e., *direct retrieval suppression* or *thought substitution*) or that leave it to the participants to prevent retrieval anyway they see fit (i.e., *unspecified* instructions). We hypothesized that participants would benefit from instructions that induce a specific mechanism. Second, we examined whether the valence of the memories influences SIF, and whether this is especially the case for participants affected by clinical/sub-clinical conditions. This is based on the idea that mood-congruent recall effects might modulate the effectiveness of memory control (Gaddy & Ingram, 2014; Matt, Vázquez, & Campbell, 1992). Third, we tested whether more repetitions of a given *suppress* cue are associated with stronger SIF, as more *repetitions* provide more opportunities for successful suppression (as suggested by, e.g., Anderson & Green, 2001; Joormann et al., 2009). Fourth, we assessed the effects of presentation time for *suppress* cues. With longer presentation times, the suppression effort has to be sustained for a more extended period. This has recently been shown to cause more memory intrusions (van Schie & Anderson, 2018). We examine whether it also reduces SIF. Fifth, to inform future developments, we explored whether the effect size of SIF is sensitive to the type of *material* that had to be suppressed (i.e., words or pictures).

# 2. Methods

## 2.1. Search strategy and inclusion criteria

We sought to identify all studies that had used the *Think/No-Think* procedure to compare healthy groups with clinical or sub-clinical samples typically associated with cognitive control difficulties. We conducted our search in PubMed, Web of Science, and Google Scholar (on September 15, 2017) using combinations of the following search terms: *Think-No Think* and/or *motivated forgetting*, and disorders-related keywords: such as *thought control ability, impulsivity*, *anxiety*, *depression*, *dysphoria*, *ADHD* (attention deficit hyyperactivity disorder)*, OCD* (obsessive-compulsive disorder), *PTSD*, *schizophrenia*, *rumination*, *addiction*, *substance abuse*, *borderline*, *repressive coping.* (The term *suppression-induced forgetting* produced consistently redundant results and was dropped from the search strategy).

Our literature search also included key terms related to questionnaires and tasks commonly associated with the broader literature on anxiety, depression, and thought control deficits. Specifically, these were the *STAI* (State-Trait Anxiety Inventory), *PANAS* (Positive and Negative Affect Schedule), *Beck Anxiety Inventory*, *Beck Depression Inventory*, *White Bear Suppression Inventory*, and the *Thought Control Ability questionnaire* (*TCAQ*; Luciano, Algarabel, Tomás, & Martínez, 2005). In addition, we consulted two recent review articles for additional references (Hulbert, Hirschstein, Brontë, & Broughton, 2018; Nørby, 2018), and included a study that was published after the initial literature search had been completed (Waldhauser et al., 2018). For exploratory purposes, we also included terms related to control and control deficits more broadly, i.e., *Stop-Signal Task*, *N-Back*, *OSPAN* (Operation Span), *BIS-11* (Barratt Impulsiveness Scale), *Rumination Response Scale*, *Go/No-Go*, *Stroop*, and *Flanker*.

The inclusion procedure for the retrieved studies is summarized in Figure 2, following the recommendation of Moher, Liberati, Tetzlaff, and Altman (2009). We included all studies that compared at least one clinical sample to a healthy control group. We also included, as sub-clinical samples, studies with groups of participants that scored high on questionnaires of clinical relevance (i.e., *BDI*, *STAI*, and *RRS*), or studies that split their participants into sub-clinical and control groups based on such questionnaires. We included only studies that were published in English (but not limited to studies that

employed English linguistic stimuli); that were peer-reviewed; that reported at least one test outcome pertaining to episodic memory performance; that reported sufficient data for the meta-analysis either in text, figures, or supplementary material; and that used the *Think/No-Think* procedure.

## 2.2. Data extraction

In total, the search yielded 208 unique entries, of which 20 entered our quantitative analysis (Table 1). For each included study, we recorded the magnitude of SIF within each group and that of their difference, the suppression mechanism induced by the *instructions*, the valence of the *suppress* items, the *type* of stimuli (words or pictures), the *repetitions* of each *suppress* item in the think-no/think phase, the presentation time for each *suppress* item (*duration*) and information about the clinical or sub-clinical condition of the target sample.

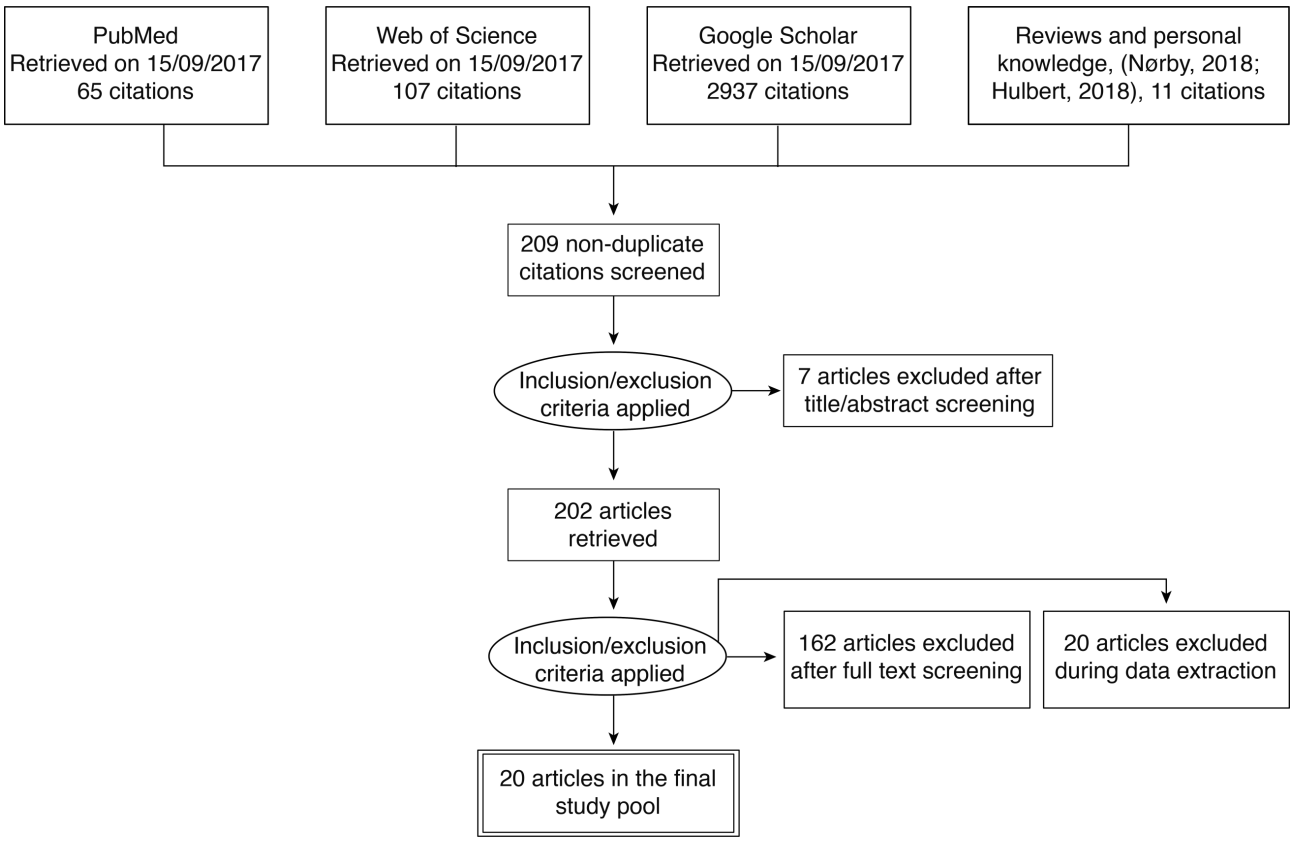


**Figure 2. Schematic overview of the literature search and inclusion process.**

Several studies reported multiple, non-independent measures of SIF (e.g., retrieval accuracy and reaction times), performance on different test formats, or multiple ways of rating the quality of the retrieved memories. Similarly, some studies employed within-participant manipulations of task features such as the number of repetitions during the *Think/No-Think* phase or the valence of the *suppress* items. They therefore provided multiple estimates of SIF (i.e., one for each level of the within-subject manipulation). In these cases, we included only one independent effect of each study. In particular, we selected the single effect that we deemed most similar to the outcome measure reported in the other included studies as well as to the standard measures in the extant *Think/No-Think* literature (see Table 1 for details). However, whenever task features were manipulated between-groups, we included all of the independent SIF effects, as long as it was possible to distinguish performance of healthy participants from that of clinical and sub-clinical samples. For example, when different groups performed the *Think/No-Think* phase with either positive or negative *suppress* items, we included in our analyses both of the resulting independent effect sizes.

For many of the included studies, the critical mean values and measures of dispersion were only provided in plots (Table 1). In these cases, we manually extracted these values using WebPlotDigitizer (Rohatgi, 2017), which has been shown to yield high inter-coder reliability (Drevon, Fulsa, and Malcolm, 2017). Each plot was coded by two of the authors (KR and DFS). When in doubt about the precise value (i.e., at the first decimal place), the respective values were averaged. We calculated the standard deviations from their respective standard errors when only the latter were available in text or plots. Lastly, the magnitude of the SIF effect was always coded such that positive values reflected greater SIF; this required that we sometimes multiply the reported value by -1 when SIF had been calculated in a reverse fashion (i.e., *suppress* – *baseline* rather than *baseline* – *suppress*; Depue, Burgess, Willcutt, Ruzic, & Banich, 2010).

In addition, we coded for the five potential moderators of SIF. First, we coded the nature of the *instructions* given to participants to prevent retrieval (*direct retrieval suppression*, *thought substitution*, or *unspecified*). One study had different participants assigned to either *unspecified* or *thought substitution* instructions (Noreen & Ridout, 2016a), but did not provide separated SIF results as a function of both, *instructions* and *group*. For each group, we therefore took the SIF effects combined across the two instruction

conditions and marked them as *unspecified*. Second, we coded the *valence* of the stimulus material as either neutral, positive, negative, or mixed (i.e., when the only reported effect sizes were combined across different valence levels). When studies comprehensively reported SIF for different valence categories assigned to the same participants (e.g., for neutral, negative, and positive memories in Marzi, Regina, & Righi, 2014; neutral and negative in Sacchett et al., 2017; Zhang, Xie, Liu, & Luo, 2016), we generally included the effect size related to neutral items. Only for one study (Dieler, Herrmann, & Fallgatter, 2014) did we code SIF for negative rather than neutral items, because its analysis of group differences (low vs. high anxiety) was based on negative items only. Third, we coded the *repetitions* of *suppress* items, i.e., the number of times that participants encountered each cue in the *Think/No-Think* phase. One study reported a SIF effect averaged across two conditions with two and eight repetitions (Noreen & Ridout, 2016a). We here coded the average (five) as the number of repetitions associated with that effect size. Fourth, we coded the duration for which cues remained on the screen during the *Think/No-Think* phase, and, fifth, the *type* of stimuli that participants had to suppress (i.e., words or pictorial material).

---

---- Table 1 about here ----

---

### 2.3. Statistical analysis

Our main focus was twofold: assessing the reliability and magnitude of SIF in healthy individuals and determining whether SIF is indeed reduced in clinical/sub-clinical samples characterized by intrusive thoughts and deficits of cognitive control. We therefore computed a series of meta-analyses in R 3.5.1 (R Development Core Team, 2008) with the package *metafor* 2.0 (Viechtbauer, 2010).

We first performed a random-effects meta-analysis (Hedges & Olkin, 1985) of the SIF effect sizes reported for the healthy control samples ($N$=534; 27 effect sizes). For each study, we computed the standardized mean change with change score standardization (as implemented in the *escalc* function; measure set to *SMCC*) based on the extracted means and standard deviations of *suppress* and *baseline* items. However, this method requires an estimate of the correlation between *baseline* and *suppress* items, which was

not reported in the surveyed literature. We thus estimated the correlation, based on data from our group, as $r$ = .4. In addition, we performed a sensitivity analysis to assess the extent to which the meta-analytical findings were influenced by the choice of correlation coefficient (see 3.1.2). Specifically, we performed two additional analyses with $r$ = .2 and .6 respectively. The second meta-analysis examined the effect sizes from clinical and sub-clinical samples ($N$ = 534; 27 effect sizes), and was based on the identical analytical approach.

We complemented the meta-analyses of the healthy and clinical/sub-clinical samples with a series of moderator analyses. We performed separate analyses for each of the five moderators (*instructions*, *valence*, *repetitions*, *duration*, and *material*) due to the relatively small pool of effect sizes. For the same reason, we always applied the Knapp and Hartung method (Knapp & Hartung, 2003) to mitigate the chance of type I error. For the *valence* models, we only retained effect sizes coded as *neutral* or *negative*, since fewer than three studies or four effect sizes contributed to the other levels of the factor (i.e., *positive* and *mixed* valence).

We further assessed whether any of the single-moderator models exhibited a better fit to the data than the simple model without moderators. Specifically, we used Akaike's Information Criterion (*AIC*; Akaike, 1998) with small-sample correction (*AICc*), transformed to conditional probabilities for each model (Wagenmakers & Farrell, 2004). The resulting *AIC* weights (*AICw*) thus provide evidence for the relative fit of the two compared models to the data (note that all *AICw* for a set of models sum up to 1). We computed *AICc* and *AICw* using the *fitstats* (from the *metafor* package) and *akaike.weights* (from the *qpcR* package; Spiess, 2018) functions in *R*. Because the *valence* model was computed on a reduced data-set, its *AICw* was compared with that of a similarly reduced simple model.

The final meta-analysis compared SIF in the healthy versus clinical/sub-clinical samples. This analysis was based on the standardized mean SIF difference (*SMD*, i.e., Hedges'*g*; Hedges, 1981) of the samples in the individual studies. We computed the *SMD* with the *escalc* function.

We followed-up on the results with an additional meta-regression that clustered studies based on clinical and sub-clinical conditions. Specifically, we further grouped the clinical

315 samples with respect to the psychiatric taxonomy of the DSM (4th ed., text rev.; DSM–IV–TR; American Psychiatric Association, 2000) and added the sub-clinical samples according to their relatedness along the psycho-pathological continuum (e.g., we combined depressed mood with major depressive disorder) (Table 1). We thus identified a *depression cluster* (15 effect sizes, including major depressive disorder, dysphoria, 320 and rumination, *N*=286 clinical/sub-clinical participants) and an *anxiety cluster* (5 effect sizes, including high trait anxiety, generalized anxiety disorder (GAD), and PTSD, *N*=90 clinical/sub-clinical participants). We further assigned the remaining effect sizes to a *mixed cluster* (6 effect sizes, with N=158 clinical/sub-clinical participants). This *cluster* included one study each on alcohol abuse, ADHD, schizophrenia, repressive coping style 325 (as measured by the Index of Self-Regulation of Emotion, *ISE*; Mendolia, 2002), low thought control ability (as measured by the *TCAQ*; Luciano, Algarabel, Tomás, & Martínez, 2005; greater scores on the *TCAQ* are negatively associated with both anxiety and depression as well as obsessive-compulsive disorder; Williams et al., 2010), and dissociative disorders (as measured by the Dissociative Experiences Scale, *DES*; 330 Bernstein & Putnam, 1986).

Meta-analyses are susceptible to publication bias, i.e., the inflation or otherwise distortion of effect-size estimates due to selective reporting of favorable study outcomes (Thornton & Lee, 2000) and other forms of questionable research practices (Renkewitz & Keiner, preprint). In the context of the present meta-analyses, publication bias might 335 take two distinct forms: First, there could be a bias for reporting experiments that yielded a significant SIF effect for the healthy individuals. Secondly, there could be a bias for publishing studies that found a significant between-group difference (i.e., SIF greater in healthy than clinical sample).

To gauge these biases, we used contour-enhanced funnel plots to display each study's 340 effect size against its precision as indexed by the standard error (Peters, Sutton, Jones, Abrams, & Rushton, 2008) (Figures 4 and 6). Compared to a traditional funnel plot, a contour-enhanced funnel plot is centered at zero, and displays areas of statistical significance. This, in turn, allows for easier visual detection of publication bias due to exclusion of studies that yielded non-significant results. We then used Egger´s 345 regression test (e.g., Peters, Sutton, Jones, Abrams, & Rushton, 2006; Egger, Smith, Schneider, & Minder, 1997) to assess funnel plot asymmetry as a formal indicator of

publication bias (with p < 0.1 as the critical value, following recommendation by Egger et al., 1997). We also applied the trim-and-fill procedure (Duval & Tweedie, 2000), which estimates the number of missing studies in the meta-analytic model due to publication bias and the impact that they might have on the meta-analytic effect size.

In cases of high heterogeneity of the estimated model, we evaluated the included effect sizes for influential cases based on a set of leave-one-out diagnostic measures (using the *influence* function in *R*) (Cook and Weisberg, 1982; Viechtbauer and Cheung, 2010). This set of measures includes externally studentized residuals (*rstudent*), difference in fits values (*dffits*; the number of standard deviations that a fitted value changes after the removal of an effect size from the model), Cook's distances (*cook.d*; how much the average effect size changes after an effect size is removed from the model), change in variance-covariance matrix of the parameter estimates (*cov.r*) after removing an effect size from the model, $\tau^2$ variance component (*tau2.del*; change in residual heterogeneity after removing an effect size from the model), Q-Statistics (*QE.del*; used to test heterogeneity after removing an effect size from the model), hat values (*hat*; indicating the leverage of each effect size in the model), and weight (*weight*; a measure of each effect size's influence).

### 2.4. Availability of data and analysis code

The data and R analysis script are openly accessible at the *Open Science Framework* (https://osf.io/f89ur/?view_only=92adb0aea0b944e196ca7d58c186da9a).

## 3. Results

### 3.1. Reliable suppression-induced forgetting in healthy individuals

Across the healthy control groups (27 effect sizes), the mean standardized difference between *baseline* and *suppress* items, i.e., the SIF effect, was 0.31, with 95% CI [0.16, 0.45], ($p$ < .001). We thus obtained evidence for a significant, small-to-moderate effect size (Figure 3A). To evaluate heterogeneity across samples, we calculated the 95% Prediction Interval (PI; IntHout, Ioannidis, Rovers, & Goeman, 2016). The PI indexes

the range of effects expected from new samples similar to those included in the analysis. This interval was broad – ranging from -0.23 to 0.85 – indicating a rather uncertain estimate. This was corroborated by a moderate amount of heterogeneity measured in terms of $I^2$ = 53.43%, 95% CI [27.68, 79.25] (Higgins, Thompson, Deeks, and Altman, 2003).

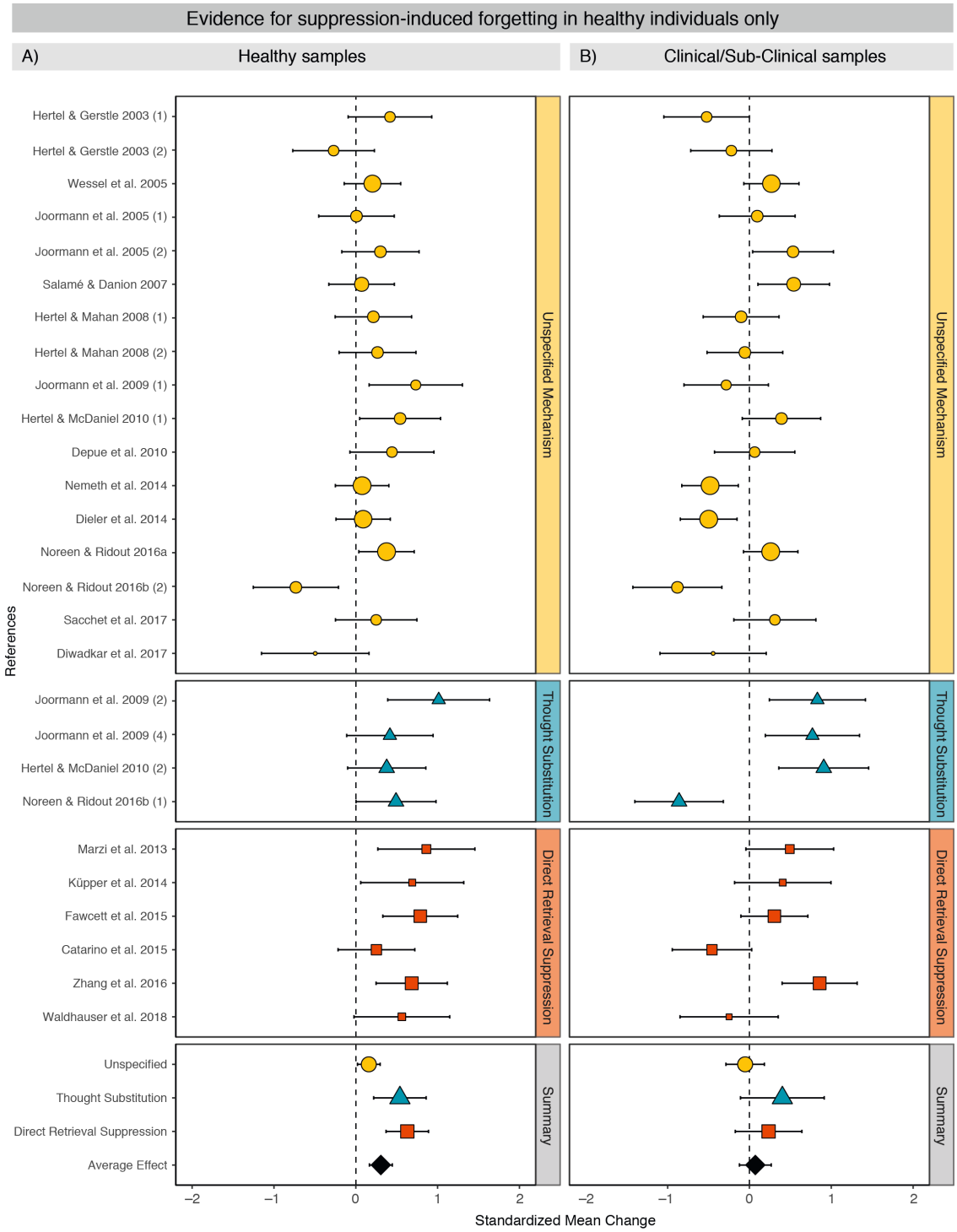

**Figure 3., Significant suppression-induced forgetting in healthy individuals only, partly moderated by instructed suppression mechanism.** Standardized mean changes with change score standardization and 95% confidence interval, separately for Panel A) healthy and Panel B) clinical/subclinical samples and as a function of the induced suppression mechanism. Symbols for individual effect sizes are sized proportionally to the respective sample sizes. Yellow circles for *unspecific*, teal triangles for *thought substitution*, and red squares for *direct retrieval suppression.* Symbols at the bottom display the meta-analytic effect sizes from the meta-regression models and the overall effect size from the random-effects model.

### *3.1.1. Evidence for suppression-induced forgetting in healthy individuals after adjustment for publication bias*

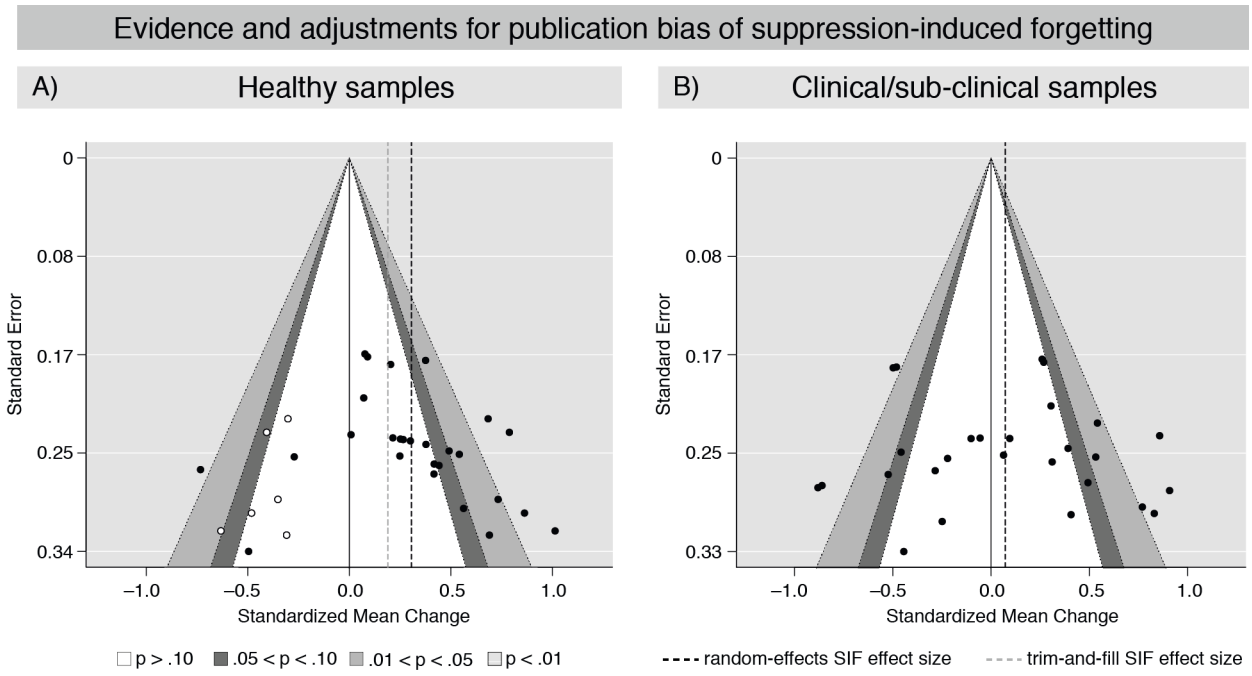


**Figure 4., Contour-enhanced funnel plots** of Panel A) healthy control and Panel B) sub-clinical samples displaying individual effect sizes (black circles). The trim-and-fill procedure added six additional data points (white circles) to achieve symmetry for the healthy samples and no additional data points for the clinical/sub-clinical samples. The black dashed lines indicate the original estimated effect sizes, whereas the gray dashed lines mark the estimated effect sizes of the trim-and-fill analyses.

We next examined the degree to which the SIF estimate is likely influenced by publication bias. For this meta-analysis, Egger´s regression test did not suggest significant publication bias, $t(25) = 1.11$, $p = .279$. By contrast, the trim-and-fill procedure estimated that six studies were missing that would be located in the area of

non-significance (Figure 4A), thus suggesting some degree of publication bias. However, adjusting for this possible bias, the newly estimated meta-analytic effect size remained significant at *SMCC* = 0.19, with 95% CI [0.04, 0.33], 95% PI [-0.50, 0.87], $p$ = .011.

### *3.1.2. Sensitivity analysis for suppression-induced forgetting in healthy control groups*

For most of the studies, we had to estimate the effect size and thus had to assume a specific correlation coefficient between *suppress* and *baseline* items. To gauge the sensitivity of the meta-analytical results to this choice (of $r$ = .4), we refitted the model two more times, with assumed $r$ values of .2 and .6.

Compared to our initial estimate, the model with $r$ = .2 yielded a somewhat smaller but significant SIF effect of 0.27, 95% CI [0.14, 0.40], 95% PI [-0.14, 0.68], $p$ < .001, whereas the model with $r$ = .6 yielded a slightly larger and significant SIF effect of 0.37, 95% CI [0.19, 0.54], 95% PI [-37, 1.10], p < 001. Heterogeneity was higher for the latter model, with $I^2$ = 67.33%, 95% CI [48.82, 85.16], compared to the former model, with $I^2$ = 39.27%, 95% CI [6.73, 73.40].

After trim-and-fill, which added six missing studies on the left side irrespective of the assumed $r$ coefficient, both models still yielded a significant SIF effect, with *SMCC* = 0.17, 95% CI [0.04, 0.29], 95% PI [-0.39, 0.73], $p$ = .011, for the $r$ = .2 model, and *SMCC* = 0.23, 95% CI [0.05, 0.40], 95% PI [-0.65, 1.10], $p$ = .011 for the $r$ = .6 model. In summary, the results are very similar for a range of assumed correlation values and thus do not seem to hinge on our particular choice.

### *3.1.3. Greater suppression-induced forgetting following direct retrieval suppression*

Though the meta-analysis provided evidence for reliable SIF in the general population, the included studies varied widely in the mechanism that individuals were instructed to adopt to prevent unwanted retrievals. Indeed, a model including *instructions* as a moderator exhibited an overwhelmingly better goodness of fit than the simple model

without a moderator, with the former ($AICw_{instructions}$ = .93) being around 13 times more plausible than the latter ($AICw_{simple}$ = .07). Consistent with this analysis, the *instructions* model (Figure 3A) significantly differentiated between the different memory control mechanisms, $F(2,24) = 12.86$, $p < .001$, displayed less residual heterogeneity $I^2$ = 27.02% 95% CI [0, 71.26], and explained greater variance in the random effect compared to the simple model, $R^2$ = 67.94. This approach thus corroborates the importance of instructing a specific mechanism to elicit SIF.

With respect to the specific mechanisms, *direct retrieval suppression* displayed a significant medium SIF effect of 0.63, with 95% CI [0.36, 0.90], 95% PI [0.23, 1.03], $p < .001$. The SIF effects for *thought substitution* and *unspecified instructions* were also significant (*thought substitution*: SIF of 0.54, 95% CI [0.20, 0.88], 95% PI [.09, .99], $p = .003$; *unspecified instructions*: SIF of 0.16, 95% CI [0.01, 0.30], 95% PI [-0.17, 0.49], $p = .033$).

Notably, the SIF effect for *direct retrieval suppression* was significantly higher than the one for *unspecified instructions* (-0.47, 95% CI [-0.78, -0.16], $p = .004$). This was also the case for the comparison of *thought substitution* and *unspecified instructions* (-0.38, 95% CI [-0.75, -0.01], $p = 0.04$). The difference between *direct retrieval suppression* and *thought substitution* was not significant, although the effect was numerically larger for the former (-0.09, 95% CI [-0.52, 0.34], $p = .671$; Figure 3A). Overall, the results thus indicate that the effectiveness of memory suppression varies with the induced mechanism.

Concerning the other moderators, the *repetitions* model ($I^2$ = 53.41%, 95% CI [27.45, 79.81]) neither provided evidence of greater SIF following a greater number of suppression attempts, $F(1, 25) = 0.55$, $p = 0.47$, nor contributed to explaining any variance, $R^2$ = 0%. In fact, model comparison yielded moderate to strong support for the simple model ($AICw_{repetitions}$ = .24, $AICw_{simple}$ = .76).

The *material* model ($I^2$ = 52.42%, 95% CI [24.76, 78.48]) revealed significant SIF for both *words* (*SMCC* = 0.25, 95% CI [0.08, 0.42], 95% PI [-0.29, 0.79], $p = .005$) and *pictures* (*SMCC* = 0.47, 95% CI [0.18, 0.76], 95% PI [-0.12, 1.06], $p = .002$). There was a numerical, but non-significant advantage of pictorial stimuli over verbal material (0.22, 95% CI [-0.11, 0.56], $p = .180$). Indeed, the *material* model also explained little

heterogeneity compared to the *simple* model, $R^2$ = 3.31%, and performed worse in model comparison ($AICw_{material}$ = .40, $AICw_{simple}$ = .60) although by a small margin (i.e., it was only 1.5 times less likely).

The *duration* model ($I^2$ = 49.93%, 95% CI [21.71, 78.13]) exhibited numerically worse SIF with longer presentation time of the *suppress* cues (-0.26, 95% CI [-0.59, 0.08], $p$ = .128). The model also explained some of the heterogeneity apparent in the *simple* model, $R^2$ = 12.44%. However, there was no clear winner when comparing the two ($AICw_{duration}$ = .52, $AICw_{simple}$ = .48).

Finally, for the *valence* model ($I^2$ = 36.81%, 95% CI [0, 73.25) we found a significant SIF effect for both neutral (*SMCC* = 0.29, 95% CI [0.12, 0.46], 95% PI [-0.12, 0.70], $p$ = .002) and negative (*SMCC* = 0.43, 95% CI [0.22, 0.65], 95% PI [0.01, 0.86], $p$ < .001) memories. Even though negative targets exhibited the greatest SIF effect numerically, this was not significantly different from SIF for neutral stimuli (0.15, 95% CI [-0.12, 0.42], $p$ = .272). Indeed, the *simple* (reduced, see 2.3) model fared better than the *valence* model in terms of information criterion ($AICw_{valence}$ = .28, $AICw_{simple}$ = .72).

### 3.2. No evidence for suppression-induced forgetting in clinical/sub-clinical samples

Having established reliable SIF in the control groups, we here turn to the corresponding effects of the clinical and sub-clinical samples. For these samples (27 effect sizes), the mean standardized difference between baseline and suppress items, i.e., the SIF effect, was 0.07, with 95% CI [-0.13, 0.28], 95% PI [-0.86, 1.01], $p$ = .474. We thus observed no evidence for SIF in these populations (Figure 3B).

However, the effect sizes exhibited a high amount of heterogeneity, as shown by the extremely wide prediction intervals and a total heterogeneity of $I^2$ = 77.58%, 95% CI [63.74, 88.55]. To test whether particular studies had driven such high heterogeneity, we assessed their individual contribution with a set of common metrics provided by the *influence* function in *R* (see 2.3.). However, we did not identify any case that significantly deviated on any of the measures.

### *3.2.1. No evidence of publication bias for suppression-induced forgetting in clinical/sub-clinical samples*

Egger´s regression test did not provide evidence for publication bias ($t(25) = 0.21$, $p = .837$), consistent with the impression from the contour enhanced funnel plot that did not suggest major departures from symmetry. Indeed, trim-and-fill did not estimate any missing studies in the opposite direction of the expected outcome, i.e., in the area of significant SIF (Figure 4B).

### *3.2.2. Sensitivity analysis for suppression-induced forgetting in clinical/sub-clinical samples*

Adopting the same rationale as in 3.1.2., we refit the simple model with either of the two alternative assumed correlation coefficients (i.e., $r = .2$ or .6). Neither of the models deviated substantially from what we had observed in the previous analysis, both in terms of the point estimate and its precision (model with $r = .2$: *SMCC* = 0.06, 95% CI [-0.11, 0.24], 95% PI [-0.73, 0.86], $p = .467$; model with $r = .6$: *SMCC* = 0.08, 95% CI [-0.16, 0.33], 95% PI [-1.07, 1.24], $p = .485$). For both models, trim-and-fill did not indicate any missing studies. Heterogeneity was also high for both models, with $I^2 = 71.68\%$, 95% CI [54.12, 85.53], for $r = .2$, and $I^2 = 83.64\%$, 95% CI [73.60, 91.66], for $r = .6$. In summary, we obtained nearly identical results irrespective of the assumed correlation value.

### *3.2.3. Clinical/sub-clinical samples show numerically greater suppression following thought substitution*

We next set out to test whether any moderators could account for the heterogeneity in the results of the clinical/sub-clinical samples. In particular, the *instructions* model only modestly improved model fit, $R^2 = 5.80\%$, and exhibited similarly high heterogeneity as the simple model ($I^2 = 76.70\%$, 95% CI [61.54, 88.39]). Consistently, the model comparison revealed better fit for the *simple* than the *instructions* model ($AICw_{instructions} = .34$, $AICw_{simple} = .66$).

However, similar to the analogous results for healthy controls, *unspecified instructions* led to the lowest SIF (Figure 3B), with *SMCC* = -0.05, 95% CI [-0.30, 0.20], 95% PI [-0.98, 0.87], $p$ = .666. *Direct retrieval suppression* instructions also did not lead to a significant SIF effect, with *SMCC* = 0.23, 95% CI [-0.19, 0.66], 95% PI [-0.75, 1.22], $p$ = .271. *Thought substitution was* associated with the largest, though not significant SIF, with *SMCC* = 0.40, 95% CI [ -0.14, 0.94], 95% PI [-0.64, .1.44], $p$ = .136. However, we did not observe a significant difference in SIF for *thought substitution* versus either *unspecified instructions* (-0.45, 95% CI [-1.05, 0.14], $p$ = .127) or versus *direct retrieval suppression* (-0.168, 95% CI [-0.86, 0.52], p = .618).

None of the other moderator models (with either *repetitions*, *duration*, *material*, or *valence* as single moderators) improved the fit in terms of $R^2$ or heterogeneity (all $I^2$´*s* > 72%). As before (3.1.5) we also compared the fit of all moderator models against that of the simple model in a pairwise fashion using *AICw* (for *valence*, compared to a reduced *simple* model). The simple model always turned out as the favored (all $AICw_{simple} > .71$).

### 3.3. Significant difference between healthy control and clinical/sub-clinical samples

Though only healthy groups exhibited significant SIF across studies, we also wanted to more directly test for an impairment of the clinical/sub-clinical samples. We therefore performed a meta-analysis of the individual studies' respective group differences. As predicted, this analysis (of 27 effect sizes) corroborated that the control samples exhibited greater SIF than their matched clinical/sub-clinical samples, with a significant standardized mean difference of 0.26 (95% CI [0.06, 0.47], $p$ = .013) (Figures 5, 6, and 7). The effect sizes were quite heterogeneous, as highlighted by a wide 95% PI [-0.56, 1.09] and a moderate total heterogeneity of $I^2$ = 58.59%, 95% CI [33.06, 79.07]. We also explored a possible source of this heterogeneity by computing influence measures. However, this approach did not detect any study that deviated from the rest of the effect sizes pool.

### *3.3.1. No evidence for publication bias for the difference of healthy vs. clinical/sub-clinical samples*

For the differences in SIF between healthy and clinical/sub-clinical samples, neither Egger´s regression test ($t(25) = 0.628$, $p = 0.536$) nor trim-and-fill provided evidence for publication bias (Figure 5). Indeed, visual inspection of the plot revealed only a slight skewness towards the presumably desired outcome.

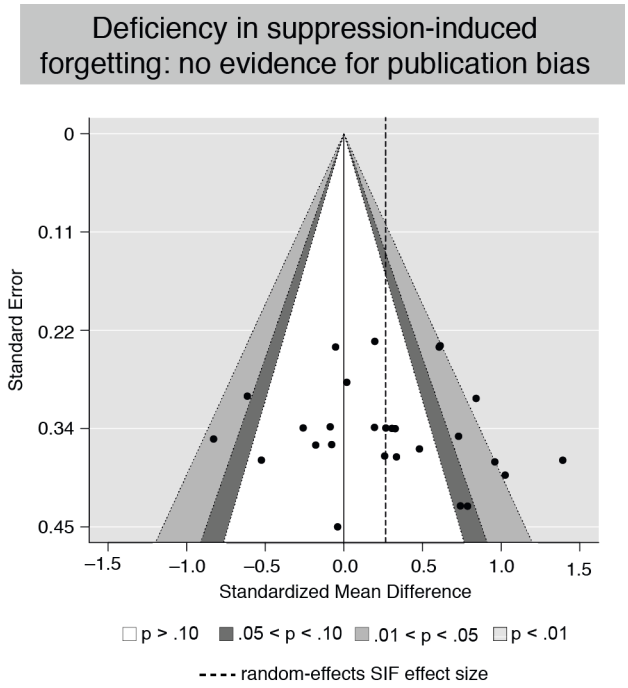


**Figure 5. Contour-enhanced funnel plot of the difference between healthy and clinical/sub-clinical samples.** Funnel plot of the effect sizes (black circles) estimated from the random-effects meta-analysis; trim-and fill did not detect any missing studies. The black dashed line indicates the original estimated effect size (*SMD* = 0.27).

### *3.3.2. Comparison of healthy vs. clinical/sub-clinical samples by clinical clusters*

Our analysis of group differences combined populations with various clinical/sub-clinical characteristics. We therefore had expected a good deal of heterogeneity. To examine whether specific conditions indeed systematically vary on SIF, we carried out a meta-regression on 27 effect sizes using the *cluster* factor described in 2.3. (i.e., *anxiety*, *depression*, and *mixed* conditions).

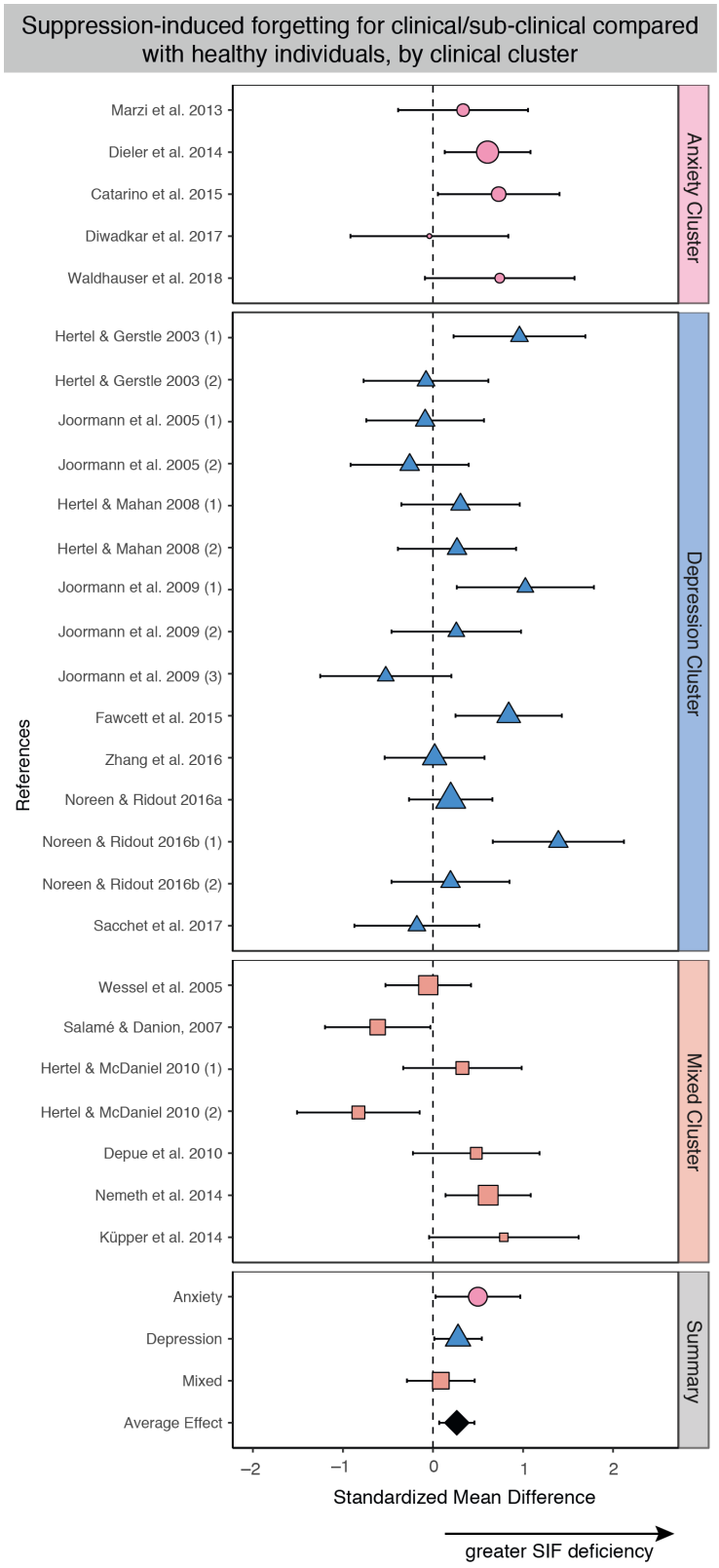


**Figure 6. Deficient suppression-induced forgetting in anxiety and depression.** Symbols indicate the standardized mean difference and 95% confidence interval between each clinical/sub-clinical sample and its respective healthy control group. Effect sizes are grouped by clinical clusters: pink circles for *anxiety*,

blue triangles for *depression*, and sand squares for *mixed*. Symbols are sized proportionally to their respective sample sizes. Symbols at the bottom display the meta-analytic effect sizes and 95% confidence intervals for each *cluster* from the respective meta-regression models plus the overall effect size from the random-effects model.

The resulting model did neither improve the heterogeneity across studies ($I^2$ = 57.97%, 95% CI [31.43, 79.75]) nor the model fit ($R^2$ = 0.99%). Furthermore, the test of the moderator was not significant $F(2, 24) = 0.91$, $p = .415$, and the model comparison provided substantial evidence in favor of the simple model compared to this moderator model ($AICw_{simple} = .79$, $AICw_{cluster} = .21$). The results thus provide no evidence that the three clusters differ from each other (Figure 6).

However, the *anxiety cluster* (including participants suffering from either PTSD, GAD, or elevated anxiety) exhibited the largest deficiency in SIF with a medium significant effect, *SMD* = 0.50, 95% CI [0, 0.99], 95 % PI [-0.44, 1.43], $p = .048$. For the *depression cluster*, we found a small significant difference, *SMD* = 0.28, 95% CI [0, 0.56], 95% PI [-0.56, 1.20], $p = .0495$. Lastly, there was no evidence for differences in SIF in the *mixed cluster*, *SMD* = 0.09, 96% CI [-0.31, 0.48], 95% PI [-0.84, 1.01], $p = .656$. Comparison of SIF difference between clusters did not yield any significant finding (all $p > .192$).

### *3.3.3. Comparison of healthy vs. clinical/sub-clinical samples by instructed suppression mechanism*

Finally, we computed an additional meta-regression for the *instructions* moderator, given that the clinical/sub-clinical samples exhibited the strongest (albeit non-significant) SIF for thought substitution, and given that this mechanism has been argued to be particularly efficient in depressed people (Joorman et al., 2009).

The resulting model, based on the 27 effect sizes, did not decrease heterogeneity across studies ($I^2$ = 57.93%, 95% CI [30.65, 79.22]), and improved model fit only slightly ($R^2$ = 3.14%). Consistently, the model comparisons also displayed worse fit for this moderator compared to the simple model ($AICw_{instructions} = .28$, $AICw_{simple} = .72$). The test of moderators was not significant, $F(2, 24) = 1.30$, $p = 0.291$, which does not provide support for differences between instructions (Figure 7).

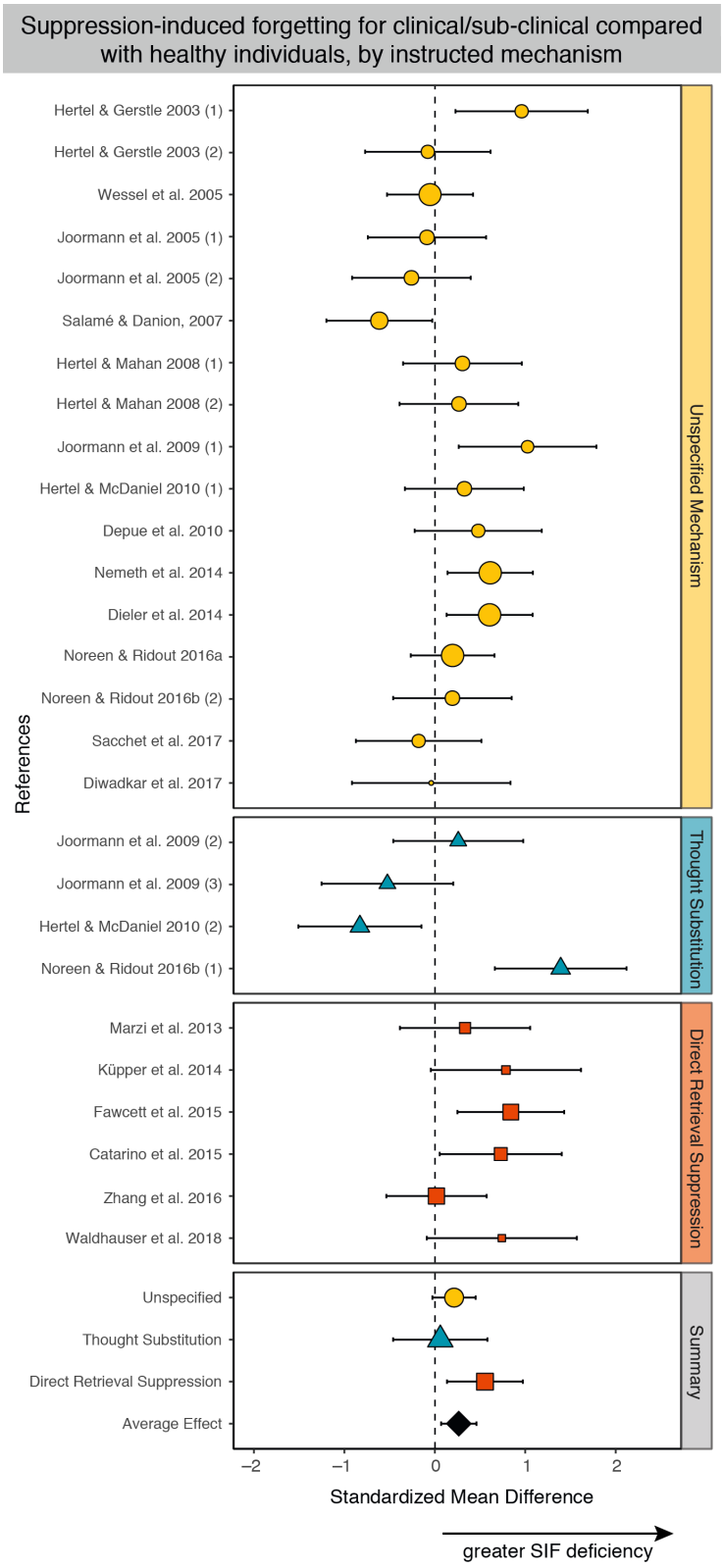


**Figure 7. Group differences in suppression-induced forgetting by induced suppression mechanism.** Standardized mean difference and 95% CI for the difference in SIF between each clinical/sub-clinical sample and its respective healthy control group. Effect sizes are grouped by *instructions*: yellow

circles for *unspecific*, teal triangles for *thought substitution*, and red squares for *direct retrieval suppression.* Symbols for individual study effect sizes are sized proportionally to their respective sample sizes. Symbols at the bottom indicate the meta-analytic effect sizes and 95% CI for each specific *instruction* from the meta-regression model and the overall SIF effect size from the random-effects model.

Nonetheless, pairwise comparisons indicated a numerically greater SIF deficiency for *direct retrieval suppression* than for either *thought substitution*, (-0.49, 95% CI [-1.20, 0.21], $p$ = .161, or *unspecified instructions* (-0.42, 95% CI [-0.85, 0.17], $p$ = .177). Furthermore, on their own, only studies instructing for *direct retrieval suppression* displayed a significant medium difference in SIF, *SMD* = 0.55, 95% CI [0.11, 1], 95% PI [-0.35, 1.46], $p$ = .016 (all other $p$ > .094).

## 4. Discussion

In a series of meta-analyses, we set to examine two questions: (i) whether preventing retrieval can cause forgetting and (ii) whether such suppression is deficient in individuals suffering from psychological disorders that are characterized by intrusive thought. We therefore focused on studies employing the *Think/No-Think* procedure that compared the SIF effect between healthy control groups and relevant clinical and sub-clinical samples. In the following, we will first discuss the replicability, effect size, mechanisms, possible causes, and moderators of SIF in the general population. We will then turn to the evidence for impaired SIF in patient populations and discuss the implications of the results for theorizing about memory suppression as a beneficial coping mechanism.

### 4.1. Reliable suppression-induced forgetting in healthy adults

Our analyses of the healthy individuals demonstrated a reliable albeit medium to small effect size. Critically, it remained significant when attempting to adjust for possible publication bias. The results thus corroborate that memory suppression can reliably induce forgetting. The healthy participants in the included studies were typically matched to the respective clinical sample on demographic measures. As a corollary, individuals of the control groups were closer to a community sample than what is usually

realized in psychology experiments. We therefore suggest that the results of this analysis may be fairly generalizable to a wider population. For the same reason, however, we might have underestimated the upper boundary of the effect size that could be achieved by high functioning, young adults. At the same time, it is also possible that we underestimated its lower boundary, because the control individuals were typically selected to lack the clinical features that had been of interest to the respective study. Therefore, the control groups may be mentally healthier than a random sample of the general population.

Across the meta-analyses, we also examined features that might influence SIF. Overall, we did not find a substantial contribution of *valence*, *repetitions*, or *duration* of suppression attempts towards the magnitude of SIF. However, the one moderator that accounted for a good part of the heterogeneity of the estimated effect were the task instructions.

Though all included studies formally used the Think/No-Think procedure to assess intentional forgetting, they differed with respect to whether they left it to the participants to find possible solutions to prevent retrieval or whether they prescribed a specific mechanism (either *direct retrieval suppression* or *thought substitution*). In healthy participants, SIF was significantly greater under *direct retrieval suppression* instructions compared to *unspecified* instructions, and numerically greater than SIF for *thought substitution*. Interestingly, at the same time, *direct retrieval suppression*, seemed to be the least effective mechanism in clinical/sub-clinical samples (see also 4.2).

These results thus clearly indicate that it is essential for future *Think/No-Think* studies to provide specific instructions. Indeed, Hertel and Calcaterra (2005) had previously provided evidence for stronger SIF when participants were using *thought substitution* rather than following *unspecified* instructions. In general, *unspecified* instructions require participants to first find possible solutions to prevent involuntary retrieval. They may also lead participants to alternate between a multitude of suppression mechanisms throughout the procedure. Both of these may diminish the efficacy of suppression and thus weaken the degree of SIF.

Furthermore, adopting specific instructions enables experimenters to identify and dissociate the precise cognitive and neural processes supporting different suppression mechanisms. For example, Benoit and Anderson (2012) provided evidence that *direct retrieval suppression* is associated with a top-down modulation of hippocampal activity that originates from the right dorsolateral prefrontal cortex. By contrast, their results indicate that *thought substitution* is based on mnemonic selection processes mediated by interactions between left ventrolateral prefrontal cortex and the hippocampus. (See also Bergström et al., 2009, for a dissociation of these mechanisms based on event-related potentials).

Different suppression mechanisms may not only differ in the underlying neuro-cognitive processes but also in the manner that they induce subsequent forgetting. The prominent inhibitory account of memory control suggests that suppression attempts lead to the recruitment of inhibitory processes that directly target and weaken the avoided memory trace (Anderson & Hanslmayr, 2014; Detre, Natarajan, Gershman, & Norman, 2013). However, in many situations, preventing retrieval may also hinder subsequent recall of the unwanted memory by non-inhibitory processes such as associative interference (Verde, 2013; Racsmány, Conway, Keresztes, & Krajcsi, A., 2012; Tomlinson, Huber, Rieth, & Davelaar, 2009; Hertel & Calcaterra, 2005). Interference may particularly contribute to forgetting following thought substitution, which likely strengthens the association between the cue (e.g., *TOMATO*) and the alternate thought or memory that participants had retrieved (e.g., *CLOWN*) to prevent the unwanted target memory from coming to mind (e.g. *VEST*).

Behavioral evidence for inhibitory versus non-inhibitory accounts of SIF is provided by studies that employed an independent probe procedure to assess forgetting (Anderson and Green, 2001; Bergström, de Fockert, & Richardson-Klavehn, 2009). In these studies, the suppressed target memories (e.g., *Asia*) are not just probed with their original cue (e.g., *clown*) – a testing procedure that would be susceptible to both interference effects and inhibition. Instead, each memory is also probed with a different cue that has a strong pre-experimental association with the memory (e.g., its category) along with a hint that uniquely points to that memory (e.g., its first letter) (e.g., *DRESS – V* for *VEST*). This test thus probes the memory while circumventing its association with the original cue (e.g., *clown*). As such, SIF on an independent probe test is unlikely

to be caused by associative interference. Instead, it is more likely caused by a weakened representation of the suppressed memory, consistent with an inhibitory account of memory control.

Of the included studies, very few had also employed an independent probe test. It is thus difficult to gauge the evidence for inhibitory versus non-inhibitory forgetting. However, an exploratory analysis of the four studies that did include such a test (comprising eight effect sizes across healthy and clinical/sub-clinical samples) revealed a trend for a small effect only, SMCC = 0.18, 95% CI [-0.0356, 0.402], p=0.088. However, all of these studies had provided *unspecified* instructions or used a *thought substitution* procedure, and particularly the latter has only inconsistently been associated with inhibitory forgetting (Bergström et al, 2009; Benoit & Anderson, 2012). These exploratory results should encourage future meta-analytical treatments of SIF as measured by independent probes, with particular attention to the instructed mechanism.

### 4.2. Compromised suppression-induced forgetting in mental disorders associated with intrusive thoughts

The meta-analysis of the non-clinical samples indicated that SIF is a replicable phenomenon in the general population. We had further hypothesized that it may constitute a beneficial coping mechanism to deal with unwanted thoughts and memories. If this were the case, we expected those individuals to be worse at suppression who find it more difficult to contain intrusive thoughts in their everyday life. To test this account, we meta-analyzed groups of participants who were either suffering from mental disorders characterized by intrusive thoughts or who were sub-clinical yet potentially susceptible to such issues as indicated by related trait measures.

In line with our hypothesis, this analysis revealed an overall negligible and non-significant SIF effect in the clinical and sub-clinical samples. Critically, their SIF was moreover reliably smaller than in the respective control samples. This deficiency was quantitatively stronger when participants tried to directly suppress the retrieval process than when they avoided the unwanted memory by retrieving a distracting substitute memory. Although in need of further validation, these results are consistent with the

theoretical assumption that thought substitution aids suppression in clinical populations (Joorman et al. 2009). At the same time, the results moreover indicate that instructions to engage in retrieval suppression may have an opposite, detrimental effect on forgetting for these individuals. This pattern may tie in with their general cognitive control deficits, given that *direct retrieval suppression* is presumably cognitively more demanding than thought substitution. Indeed, only the former mechanism is thought to be associated with top-down inhibitory control processes (e.g., Racsmány et al., 2012; Bergström et al., 2009). Therefore, while it is generally fruitful to aid suppression by prescribing a specific mechanism through precise instructions, it is also important to choose a mechanism suitable to the targeted population.

Moreover, we observed a numerically larger impairment for participants suffering from PTSD, GAD, or elevated anxiety. These data thus corroborate prior evidence from individual studies that had reported a negative association between SIF and trait anxiety (Benoit et al., 2016; Waldhauser et al., 2018), poor thought control ability (Catarino et al., 2015), depressed mood (Zhang et al., 2016), or rumination (Fawcett et al., 2015) Moreover, a similar pattern has been reported on indirect measures of memory performance (Hertel, Maydon, Ogilvie, & Mor, 2018). The pattern is also consistent with several studies that had similarly related deficient control processes at retrieval with clinical phenomena (e.g., GAD, Kircanski et al., 2016; clinical depression, Groome & Sterkaj, 2008; substance-related and addictive disorders, Stramaccia, Penolazzi, Monego, Manzan, Castelli, & Galfano, 2017). More generally, these results are consistent with a recent meta-analysis that associated broader cognitive control deficits with negative thinking (Zetsche, Bürkner, and Schulze, 2018).

The present meta-analyses focused on memory control at the stage of *retrieval*. We consider this a relevant stage for the treatment of psychological disorders, seeing that patients suffering from PTSD, for example, tend to seek help after intrusive memories would already have had time to consolidate (Marks et al., 2018). Similarly, psychological conditions such as depression and anxiety are also characterized by a problematic focus on past memories in the form of rumination (Koval, Kuppens, Allen, Sheeber, 2012; Michael, Halligan, Clark, & Ehlers, 2007). Indeed, rumination appears to be a transdiagnostic feature that is also relevant to PTSD (Birrer & Michael, 2011). Due to the delay between initial experience and subsequent treatment, it may often not be

feasible to administer potential interventions at earlier stages (see Visser et al., 2018), such as the recently proposed computer-game based intervention that is effective prior to consolidation of traumatic experiences (Iyadurai et al., 2018).

However, we note that memory control impairments in clinical populations may already manifest at the earlier stage of encoding. This has been shown for depression (Power, Dalgleish, Claudio, Tata, & Kentish, 2000; Xie, Jiang, & Zhang 2018), anxiety (Yang, Lei, & Anderson, 2010; Dieler et al., 2014), and PTSD (Cottencin et al., 2006). Overall, the recurrent finding of *memory control* deficits in these clinical populations hints at its potential transdiagnostic value (see McTeague et al., 2017).

The provided evidence for the benefits of memory suppression may constitute a conundrum when also considering clinical evidence that ties suppression to negative outcomes in trauma-related disorders (Holmes, Moulds, & Kavanagh, 2007). We certainly want to emphasize that we do not claim that it is always beneficial to try to suppress unwanted memories. In general, we believe that it is critical to engage with negative life experiences and emotions and to integrate them into who we are (see also Biglan, Hayes, & Pistorello, 2008; for examples of negative consequences of suppression in different contexts, see Le & Impett, 2016; Srivastava, Tamir, McGonigal, John, & Gross, 2009; Dalgleish & Yiend, 2006). However, we also think that there are several factors that can reconcile a beneficial take on suppression with the apparently contradicting clinical experience.

First, if the ability to suppress memories is deficient in people suffering from intrusive memories, then asking them to suppress an unwanted memory may in fact have the paradoxical effect of aggravating symptoms. That is, for them, attempts to suppress may be bound to fail and thus counterproductively induce rehearsal – and thus strengthening – of unwanted memories.

Secondly, there seems to be a difference between the notion of suppression as framed in the *Think/No-Think* literature versus the literature on cognitive-behavioral therapy and emotion regulation (Engen and Anderson, 2018). Specifically, direct retrieval suppression should not be confused with expressive suppression. Expressive suppression refers to the act of voluntarily inhibiting overt expressions of one's emotional states – for instance, facial expressions or tone of voice (Suchy, 2015) – to

one´s internal emotional states. This, in turn, has been extensively associated with poorer well-being (Haga, Kraft, & Corby, 2009; Moore, Zoellner, & Mollenholt, 2008).

Finally, as pointed out by Visser et al. (2018), it may well turn out to be critical what aspects of a memory are being targeted. In particular, these authors argue that it may be beneficial to spare the declarative component of a memory while attenuating its emotional component.

These issues notwithstanding, we propose that memory suppression can serve as a mechanism that helps us control the intrusive retrieval of unwanted memories (Anderson et al., 2004; Depue, et al., 2007; Benoit, Hulbert, Huddleston, & Anderson, 2015). Consistent with this proposal, a recent study provided evidence that, in a sample of non-clinical individuals, those were better at suppressing unwanted memories that had a greater lifetime exposure to traumatic incidents (Hulbert & Anderson, 2018). These data suggest that people employ suppression to cope with traumatic experiences and that such practice actually boosts the efficiency of this process.

### 4.3. Caveats

Importantly, due to the designs of the primary studies, we are not able to infer causal relationship between reduced SIF and psychopathology. Prospective studies are needed to disentangle whether SIF impairments precede or follow psychopathology, and to determine their potential role as a disorder-maintaining factor. In this respect, the objective difficulty in obtaining large clinical samples and adequate matched controls, combined with the relatively small effect size (at least when studies do not prescribe a specific suppression mechanism), call for joint efforts to investigate such causal relationships.

We also want to point out some limitations of the extant literature and the current meta-analyses. Based on the meta-analytical effect size, we note that the primary studies had used fairly low sample sizes overall (on average, about 20 participants per group), and that they thus were certainly low-powered to detect either within-group SIF (~0.42 power for a one-tailed test) or a between-group difference in the magnitude of SIF (~0.31 power for a one-tailed test) (as computed with R package *pwr*, Champely, 2015). Note, however, that studies prescribing either thought substitution or direct

retrieval suppression yielded considerably greater effect sizes. Furthermore, the limited number of effects in each clinical *cluster* did not allow us to conduct more fine grained analyses such as for a possible gradient of impairment associated with the severity of the disorders.

Our conclusions should thus be evaluated with respect to the low power of some of the primary data and the relatively high heterogeneity of some of the meta-analytic models (in particular concerning the clinical/sub-clinical samples). In addition, we had limited means to assess the impact of some of the chosen moderators. For this reason, it would be worthwhile to examine other aspects of the *Think/No-Think* task in future meta-analytic endeavors (such as the number of suppressed target memories, specific features of the learning procedure, and the maximum allotted time for memory retrieval on the final test).

Finally, methods including Egger's regression and trim-and-fill may underestimate the presence of publication bias, especially so in the context of highly heterogeneous models (Renkewitz & Keiner, preprint). Therefore, any meta-analytical efforts should ideally be complemented by pre-registered, large-scale replication attempts. Nonetheless, we suggest that meta-analyses currently provide the best available evidence on SIF and its disturbance in clinical disorders.

### 4.4. Conclusions

In light of the present results, we therefore suggest that SIF is the replicable hallmark of an effortful process that allows us to voluntarily prevent memory retrieval. Importantly, the observation that this ability is associated with psychological well-being indicates that it may indeed constitute an adaptive coping mechanism. We certainly neither propose that preventing retrieval is always beneficial nor do we suggest that fostering suppression would necessarily be an adequate therapeutic intervention. Yet, in our everyday life, it may help us control intrusive and unwanted thoughts and thus allow us to edit the contents of our memories.

## Role of funding sources

RGB, DFS, and KMR were funded by a Max Planck Research Group awarded to RGB. JMF was funded by an NSERC Discovery Grant.

## Acknowledgements

We are grateful to Ann-Kristin Meyer for her insightful comments on the manuscript and Kerstin Flake for her assistance in creating the figures.

# References

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of Hirotugu Akaike* (pp. 199-213). Springer, New York, NY.

Algarabel, S., Luciano, J. V., & Martínez, J. L. (2006). Inhibitory Voluntary Control of Memory: Effect of Stimulus Onset Asynchrony on Reaction Time to Suppressed Memories. *Psicologica: International Journal of Methodology and Experimental Psychology*, *27*(1), 57-77.

American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders (4th ed., text rev.). Washington, DC: Author.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Washington, DC: Author.

Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature*, *410*(6826), 366.

Anderson, M. C., Ochsner, K. N., Kuhl, B., Cooper, J., Robertson, E., Gabrieli, S. W., ... & Gabrieli, J. D. (2004). Neural systems underlying the suppression of unwanted memories. *Science*, *303*(5655), 232-235.

Anderson, M. C., & Huddleston, E. (2012). Towards a cognitive and neurobiological model of motivated forgetting. In *True and false recovered memories* (pp. 53-120). Springer, New York, NY.

Anderson, M. C., & Hanslmayr, S. (2014). Neural mechanisms of motivated forgetting. *Trends in Cognitive Sciences*, *18*(6), 279-292.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). Identifying influential data and sources of collinearity. Regression Diagnostics.

Benoit, R. G., & Anderson, M. C. (2012). Opposing mechanisms support the voluntary forgetting of unwanted memories. *Neuron*, *76*(2), 450-460.

Benoit, R. G., Hulbert, J. C., Huddleston, E., & Anderson, M. C. (2015). Adaptive top–down suppression of hippocampal activity and the purging of intrusive memories from consciousness. *Journal of Cognitive Neuroscience*, *27*(1), 96-111.

Benoit, R. G., Davies, D. J., & Anderson, M. C. (2016). Reducing future fears by suppressing the brain mechanisms underlying episodic simulation. *Proceedings of the National Academy of Sciences*, *113*(52), E8492-E8501.

Bergström, Z. M., Velmans, M., de Fockert, J., & Richardson-Klavehn, A. (2007). ERP evidence for successful voluntary avoidance of conscious recollection. Brain research, 1151, 119-133.

Bergström, Z. M., de Fockert, J. W., & Richardson-Klavehn, A. (2009). ERP and behavioural evidence for direct suppression of unwanted memories. *NeuroImage*, *48*(4), 726-737.

Bernstein, E. M., & Putnam, F. W. (1986). Development, reliability, and validity of a dissociation scale. *Journal of nervous and mental disease*.

Birrer, E., & Michael, T. (2011). Rumination in PTSD as well as in traumatized and non-traumatized depressed patients: A cross-sectional clinical study. *Behavioural and Cognitive Psychotherapy*, *39*(4), 381-397.

Brewin, C. R. (2014). Episodic memory, perceptual memory, and their interaction: foundations for a theory of posttraumatic stress disorder. *Psychological Bulletin*, *140*(1), 69.

Bulevich, J. B., Roediger, H. L., Balota, D. A., & Butler, A. C. (2006). Failures to find suppression of episodic memories in the think/no-think paradigm. *Memory & cognition*, *34*(8), 1569-1577.

Catarino, A., Küpper, C. S., Werner-Seidler, A., Dalgleish, T., & Anderson, M. C. (2015). Failing to forget: Inhibitory-control deficits compromise memory suppression in posttraumatic stress disorder. *Psychological Science*, *26*(5), 604-616.

Champely, S. (2015). pwr: Basic functions for power analysis. R package version, 1(1).

Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*. New York: Chapman and Hall.

Cottencin, O., Vaiva, G., Huron, C., Devos, P., Ducrocq, F., Jouvent, R., ... & Thomas, P. (2006). Directed forgetting in PTSD: a comparative study versus normal controls. *Journal of Psychiatric Research*, *40*(1), 70-80.

Dalgleish, T., & Yiend, J. (2006). The effects of suppressing a negative autobiographical memory on concurrent intrusions and subsequent autobiographical recall in dysphoria. Journal of Abnormal Psychology, 115(3), 467.

Depue, B. E., Curran, T., & Banich, M. T. (2007). Prefrontal regions orchestrate suppression of emotional memories via a two-phase process. S*cience*, *317*(5835), 215-219.

Depue, B. E., Burgess, G. C., Willcutt, E. G., Ruzic, L., & Banich, M. T. (2010). Inhibitory control of memory retrieval and motor processing associated with the right lateral prefrontal cortex: evidence from deficits in individuals with ADHD. *Neuropsychologia*, *48*(13), 3909-3917.

Detre, G. J., Natarajan, A., Gershman, S. J., & Norman, K. A. (2013). Moderate levels of activation lead to forgetting in the think/no-think paradigm. *Neuropsychologia*, *51*(12), 2371-2388.

Dieler, A. C., Herrmann, M. J., & Fallgatter, A. J. (2014). Voluntary suppression of thoughts is influenced by anxious and ruminative tendencies in healthy volunteers. *Memory*, *22*(3), 184-193.

Diwadkar, V. A., Re, M., Cecchetto, F., Garzitto, M., Piccin, S., Bonivento, C., ... & Brambilla, P. (2017). Attempts at memory control induce dysfunctional brain activation profiles in Generalized Anxiety Disorder: An exploratory fMRI study. *Psychiatry Research: Neuroimaging*, *266*, 42-52.

Drevon, D., Fursa, S. R., & Malcolm, A. L. (2017). Intercoder reliability and validity of WebPlotDigitizer in extracting graphed data. *Behavior Modification*, *41*(2), 323-339.

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel- plot–based method of testing and adjusting for publication bias in meta- analysis. *Biometrics*, *56*(2), 455-463.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. British Medical Journal, 315, 629–634.

Ehlers, A., Hackmann, A., & Michael, T. (2004). Intrusive re- experiencing in post-traumatic stress disorder: Phenomenology, theory, and therapy. *Memory*, *12*(4), 403-415.

Engen, H. G., & Anderson, M. C. (2018). Memory control: a fundamental mechanism of emotion regulation. *Trends in Cognitive Sciences*.

Fawcett, J. M., Benoit, R. G., Gagnepain, P., Salman, A., Bartholdy, S., Bradley, C., ... & Anderson, M. C. (2015). The origins of repetitive thought in rumination: Separating cognitive style from deficits in inhibitory control over memory. *Journal of Behavior Therapy and Experimental Psychiatry*, *47*, 1-8.

Gaddy, M. A., & Ingram, R. E. (2014). A meta-analytic review of mood-congruent implicit memory in depressed mood. *Clinical Psychology Review*, *34*(5), 402-416.

Gagnepain, P., Henson, R. N., & Anderson, M. C. (2014). Suppressing unwanted memories reduces their unconscious influence via targeted cortical inhibition. *Proceedings of the National Academy of Sciences*, 201311468.

Goschke, T. (2014). Dysfunctions of decision‐making and cognitive control as transdiagnostic mechanisms of mental disorders: advances, gaps, and needs in current research. *International Journal of Methods in Psychiatric Research*, *23*(S1), 41-57.

Groome, D., & Sterkaj, F. (2010). Retrieval-induced forgetting and clinical depression. *Cognition and Emotion*, *24*(1), 63-70.

Hackmann, A., Ehlers, A., Speckens, A., & Clark, D. M. (2004). Characteristics and content of intrusive memories in PTSD and their changes with treatment. *Journal of Traumatic Stress: Official Publication of The International Society for Traumatic Stress Studies*, *17*(3), 231-240.

Haga, S. M., Kraft, P., & Corby, E. K. (2009). Emotion regulation: Antecedents and well‐being outcomes of cognitive reappraisal and expressive suppression in cross-cultural samples. *Journal of Happiness Studies*, 10(3), 271-291.

Biglan, A., Hayes, S. C., & Pistorello, J. (2008). Acceptance and commitment: Implications for prevention science. *Prevention Science*, 9(3), 139-152.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics,* 6(2), 107-128.

Hedges, L., & Olkin, I. (1985). *Statistical models for meta-analysis*.

Hertel, P. T. (1997). On the contributions of deficient cognitive control to memory impairments in depression. *Cognition & Emotion*, *11*(5-6), 569-583.

Hertel, P. T. (1998). Relation between rumination and impaired memory in dysphoric moods. Journal of Abnormal Psychology, 107(1), 166.

Hertel, P. T., & Gerstle, M. (2003). Depressive deficits in forgetting. *Psychological Science*, *14*(6), 573-578.

Hertel, P. T., & Calcaterra, G. (2005). Intentional forgetting benefits from thought substitution. *Psychonomic Bulletin & Review*, *12*(3), 484-489.

Hertel, P. T. (2007). Impairments in inhibition or cognitive control in psychological disorders. *Applied and Preventive Psychology*, *12*(3), 149-153.

Hertel, P. T., & Mahan, A. (2008). Depression-related differences in learning and forgetting responses to unrelated cues. *Acta Psychologica*, *127*(3), 636-644.

Hertel, P., & McDaniel, L. (2010). The suppressive power of positive thinking: Aiding suppression-induced forgetting in repressive coping. *Cognition and Emotion*, *24*(7), 1239-1249.

Hertel, P. T., Maydon, A., Ogilvie, A., & Mor, N. (2018). Ruminators (Unlike Others) Fail to Show Suppression-Induced Forgetting on Indirect Measures of Memory. *Clinical Psychological Science*, *6*(6), 872-881.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal*, *327*(7414), 557.

Holmes, E. A., Moulds, M. L., & Kavanagh, D. (2007). Memory suppression in PTSD treatment?. *Science*, *318*(5857), 1722-1722.

Hulbert, J. C., & Anderson, M. C. (2018). What doesn't kill you makes you stronger: Psychological trauma and its relationship to enhanced memory control. *Journal of Experimental Psychology: General*, *147*(12), 1931.

Hulbert, J. C., Hirschstein, Z., Brontë, C. A., & Broughton, E. (2018). Unintended side effects of a spotless mind: theory and practice. *Memory*, *26*(3), 306-320.

Iyadurai, L., Blackwell, S. E., Meiser-Stedman, R., Watson, P. C., Bonsall, M. B., Geddes, J. R., ... & Holmes, E. A. (2018). Preventing intrusive memories after trauma via a brief intervention involving Tetris computer game play in the emergency department: a proof-of-concept randomized controlled trial. *Molecular psychiatry*, *23*(3), 674.

Iyadurai, L., Visser, R. M., Lau-Zhu, A., Porcheret, K., Horsch, A., Holmes, E. A., & James, E. L. (2018). Intrusive memories of trauma: A target for research

bridging cognitive science and its clinical application. *Clinical Psychology Review*.

IntHout, J., Ioannidis, J. P., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ open*, *6*(7), e010247.

James, W. (1892). *Text-book of Psychology*. Macmillan.

Joormann, J., Hertel, P. T., Brozovich, F., & Gotlib, I. H. (2005). Remembering the good, forgetting the bad: intentional forgetting of emotional material in depression. *Journal of Abnormal Psychology*, *114*(4), 640.

Joormann, J., Hertel, P. T., LeMoult, J., & Gotlib, I. H. (2009). Training forgetting of negative material in depression. *Journal of Abnormal Psychology*, *118*(1), 34.

Kircanski, K., Joormann, J., & Gotlib, I. H. (2012). Cognitive aspects of depression. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*(3), 301-313.

Kircanski, K., Johnson, D. C., Mateen, M., Bjork, R. A., & Gotlib, I. H. (2016). Impaired retrieval inhibition of threat material in generalized anxiety disorder. *Clinical Psychological Science*, *4*(2), 320-327.

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in medicine*, *22*(17), 2693-2710.

Koval, P., Kuppens, P., Allen, N. B., & Sheeber, L. (2012). Getting stuck in depression: The roles of rumination and emotional inertia. *Cognition & emotion*, *26*(8), 1412-1427.

Küpper, C. S., Benoit, R. G., Dalgleish, T., & Anderson, M. C. (2014). Direct suppression as a mechanism for controlling unpleasant memories in daily life. *Journal of Experimental Psychology: General*, *143*(4), 1443.

Le, B. M., & Impett, E. A. (2016). The costs of suppressing negative emotions and amplifying positive emotions during parental caregiving. Personality and Social Psychology Bulletin, 42(3), 323-336.

Luciano, J. V., Algarabel, S., Tomás, J. M., & Martínez, J. L. (2005). Development and validation of the thought control ability questionnaire. *Personality and Individual Differences*, *38*(5), 997-1008.

Marks, E. H., Franklin, A. R., & Zoellner, L. A. (2018). Can't get it out of my mind: A systematic review of predictors of intrusive memories of distressing events. *Psychological Bulletin*, *144*(6), 584.

Marzi, T., Regina, A., & Righi, S. (2014). Emotions shape memory suppression in trait anxiety. *Frontiers in Psychology*, *4*, 1001.

Matt, G. E., Vázquez, C., & Campbell, W. K. (1992). Mood-congruent recall of affectively toned stimuli: A meta-analytic review. *Clinical Psychology Review*, *12*(2), 227-255.

McTeague, L. M., Goodkind, M. S., & Etkin, A. (2016). Transdiagnostic impairment of cognitive control in mental illness. *Journal of Psychiatric Research*, *83*, 37-46.

Mecklinger, A., Parra, M., & Waldhauser, G. T. (2009). ERP correlates of intentional forgetting. *Brain Research*, *1255*, 132-147.

Mendolia, M. (2002). An index of self-regulation of emotion and the study of repression in social contexts that threaten or do not threaten self-concept. *Emotion*, *2*(3), 215.

Michael, T., Halligan, S. L., Clark, D. M., & Ehlers, A. (2007). Rumination in posttraumatic stress disorder. *Depression and anxiety*, *24*(5), 307-317.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine*, *151*(4), 264-269.

Moore, S. A., Zoellner, L. A., & Mollenholt, N. (2008). Are expressive suppression and cognitive reappraisal associated with stress-related symptoms? *Behaviour research and therapy*, *46*(9), 993-1000.

Nemeth, V. L., Kurgyis, E., Csifcsak, G., Maraz, A., Almasi, D. A., Drotos, G., ... & Must, A. (2014). The impact of intermediate-term alcohol abstinence on memory retrieval and suppression. *Frontiers in Psychology*, *5*, 1396.

Nørby, S. (2015). Why forget? On the adaptive value of memory loss. Perspectives on *Psychological Science*, *10*(5), 551-578.

Nørby, S. (2018). Forgetting and emotion regulation in mental health, anxiety and depression. *Memory*, *26*(3), 342-363.

Noreen, S., & Ridout, N. (2016a). Intentional forgetting in dysphoria: investigating the inhibitory effects of thought substitution using independent cues. *Journal of Behavior Therapy and Experimental Psychiatry*, *52*, 110-118.

Noreen, S., & Ridout, N. (2016b). Examining the impact of thought substitution on intentional forgetting in induced and naturally occurring dysphoria. *Psychiatry Research*, *241*, 280-288.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2006). Comparison of two methods to detect publication bias in meta-analysis. *Jama*, *295*(6), 676-680.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, *61*(10), 991-996.

Power, M. J., Dalgleish, T., Claudio, V., Tata, P., & Kentish, J. (2000). The directed forgetting task: Application to emotionally valent material. *Journal of Affective Disorders*, *57*(1-3), 147-157.

R Development Core Team (2008)

Racsmány, M., Conway, M. A., Keresztes, A., & Krajcsi, A. (2012). Inhibition and interference in the think/no-think task. *Memory & Cognition*, *40*(2), 168-176.

Renkewitz, F., & Keiner, M. (2018, December 20). How to detect publication bias in psychological research? A comparative evaluation of six statistical methods. https://doi.org/10.31234/osf.io/w94ep

Rohatgi, A. (2017). WebPlotDigitizer (version 3.11) [Web-based application].

Sacchet, M. D., Levy, B. J., Hamilton, J. P., Maksimovskiy, A., Hertel, P. T., Joormann, J., ... & Gotlib, I. H. (2017). Cognitive and neural consequences of memory suppression in major depressive disorder. *Cognitive, Affective, & Behavioral Neuroscience*, *17*(1), 77-93.

Salamé, P., & Danion, J. M. (2007). Inhibition of inappropriate responses is preserved in the think-no-think and impaired in the random number generation tasks in schizophrenia. *Journal of the International Neuropsychological Society*, *13*(2), 277-287.

van Schie, K., & Anderson, M. C. (2017). Successfully controlling intrusive memories is harder when control must be sustained. *Memory*, *25*(9), 1201-1216.

Spiess, A.-N. (2018). qpcR: Modelling and Analysis of Real-Time PCR Data. R package version 1.4-1.

Srivastava, S., Tamir, M., McGonigal, K. M., John, O. P., & Gross, J. J. (2009). The social costs of emotional suppression: A prospective study of the transition to college. Journal of personality and social psychology, 96(4), 883.

Storm, B. C. (2011). The benefit of forgetting in thinking and remembering. *Current Directions in Psychological Science*, *20*(5), 291-295.

Stramaccia, D. F., Penolazzi, B., Monego, A. L., Manzan, A., Castelli, L., & Galfano, G. (2017). Suppression of competing memories in substance-related and addictive disorders: A retrieval-induced forgetting study. *Clinical Psychological Science*, *5*(2), 410-417.

Suchy, Y. (2015). Executive functioning: A comprehensive guide for clinical practice. Oxford University Press.

Sullivan, D. R., Marx, B., Chen, M. S., Depue, B. E., Hayes, S. M., & Hayes, J. P. (2019). Behavioral and neural correlates of memory suppression in PTSD. *Journal of psychiatric research*.

Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis: its causes and consequences. *Journal of Clinical Epidemiology*, *53*(2), 207-216.

Tomlinson, T. D., Huber, D. E., Rieth, C. A., & Davelaar, E. J. (2009). An interference account of cue-independent forgetting in the no-think paradigm. *Proceedings of the National Academy of Sciences*, *106*(37), 15588-15593.

Verde, M. F. (2013). Retrieval-induced forgetting in recall: Competitor interference revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5), 1433.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3).

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta- analysis. Research synthesis methods, 1(2), 112-125.

Visser, R. M., Lau-Zhu, A., Henson, R. N., & Holmes, E. A. (2018). Multiple memory systems, multiple time points: how science can inform treatment to control the expression of unwanted emotional memories. *Phil. Trans. R. Soc. B*, *373*(1742), 20170209.

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic bulletin & review*, *11*(1), 192-196.

Waldhauser, G. T., Dahl, M. J., Ruf-Leuschner, M., Müller-Bamouh, V., Schauer, M., Axmacher, N., ... & Hanslmayr, S. (2018). The neural dynamics of deficient memory control in heavily traumatized refugees. *Scientific Reports*, *8*(1), 13132.

Wessel, I., Wetzels, S., Jelicic, M., & Merckelbach, H. (2005). Dissociation and memory suppression: A comparison of high and low dissociative individuals' performance on the think–no think task. *Personality and Individual Differences*, *39*(8), 1461-1470.

Williams, A. D., Moulds, M. L., Grisham, J. R., Gay, P., Lang, T., Kandris, E., ... & Yap, C. (2010). A psychometric evaluation of the Thought Control Ability Questionnaire (TCAQ) and the prediction of cognitive control. *Journal of Psychopathology and Behavioral Assessment*, *32*(3), 397-405.

World Health Organization. (2018). International statistical classification of diseases and related health problems (11th Revision). Retrieved from https://icd.who.int/browse11/l-m/en

Xie, H., Jiang, D., & Zhang, D. (2018). Individuals with depressive tendencies experience difficulty in forgetting negative material: two mechanisms revealed by ERP data in the directed forgetting paradigm. *Scientific Reports*, *8*(1), 1113.

Yang, T., Lei, X., & Anderson, M. (2016). Decreased inhibitory control of negative information in directed forgetting. *International Journal of Psychophysiology*, *100*, 44-51.

Zetsche, U., Bürkner, P. C., & Schulze, L. (2018). Shedding light on the association between repetitive negative thinking and deficits in cognitive control – a meta-analysis. *Clinical Psychology Review*.

Zhang, D., Xie, H., Liu, Y., & Luo, Y. (2016). Neural correlates underlying impaired memory facilitation and suppression of negative material in depression. *Scientific Reports*, *6*, 37556.

| Study | Year | Comments | Instructions (mechanism) | Stimuli | | Duration (s) | Repetitions | Clinical sample | Clinical cluster | *N* | | DV | SIF (SMCC) | | SIF (SMD) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Material | Valence | | | | | Clinical | Healthy | | Clinical | Healthy | Difference |
| Catarino et al. | 2015 | | Direct RS | Pictures | Negative | 3 | 10 | PTSD | ANX | 18 | 18 | Correct ID | -0.46 | 0.25 | 0.73 |
| Depue et al. | 2010 | | unspecified | Pictures | Negative | 3.5 | 12 | ADHD | Mixed | 16 | 16 | Retrieval ACC | 0.06 | 0.44 | 0.48 |
| Dieler et al. | 2014 | | unspecified | Pictures | Negative | 4 | 12 | High trait anxiety | ANX | 36 | 35 | Retrieval ACC | -0.50 | 0.09 | 0.61 |
| Diwadkar et al. | 2017 | | unspecified | Words | Neutral | 4 | 8 | GAD | ANX | 10 | 10 | Retrieval ACC | -0.44 | -0.50 | -0.04 |
| Fawcett et al. | 2015 | | Direct RS | Words | Neutral | 3 | 12 | High rumination | DEP | 24 | 24 | Retrieval ACC | 0.30 | 0.79 | 0.84 |
| Hertel & Gerstle (1) | 2003 | Positive cue | unspecified | Words | Neutral | 3 | 16 | Dysphoria | DEP | 16 | 16 | Retrieval ACC | -0.52 | 0.42 | 0.96 |
| Hertel & Gerstle (2) | 2003 | Negative cue | unspecified | Words | Neutral | 3 | 16 | Dysphoria | DEP | 16 | 16 | Retrieval ACC | -0.22 | -0.27 | -0.08 |
| Hertel & Mahan (1) | 2008 | Related pair | unspecified | Words | Neutral | 3 | 12 | Dysphoria | DEP | 18 | 18 | Retrieval ACC | -0.10 | 0.21 | 0.31 |
| Hertel & Mahan (2) | 2008 | Unrelated pair | unspecified | Words | Neutral | 3 | 12 | Dysphoria | DEP | 18 | 18 | Retrieval ACC | -0.06 | 0.27 | 0.27 |
| Hertel & McDaniel (1) | 2010 | Unaided Suppression | unspecified | Words | Negative | 3 | 12 | High repression | Mixed | 18 | 18 | Retrieval ACC | 0.39 | 0.54 | 0.33 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Hertel & McDaniel (2) | 2010 | Aided Suppression | Thought Sub | Words | Negative | 3 | 12 | High repression | Mixed | 18 | 18 | Retrieval ACC | 0.91 | 0.38 | -0.83 |
| Joormann et al. (1) | 2009 | Unaided substitute | unspecified | Words | Negative | 3 | 12 | Depression | DEP | 15 | 15 | Retrieval ACC | -0.28 | 0.73 | 1.03 |
| Joormann et al. (2) | 2009 | Positive substitute | Thought Sub | Words | Negative | 3 | 12 | Depression | DEP | 15 | 15 | Retrieval ACC | 0.83 | 1.01 | 0.26 |
| Joormann et al. (3) | 2009 | Negative substitute | Thought Sub | Words | Negative | 3 | 12 | Depression | DEP | 15 | 15 | Retrieval ACC | 0.77 | 0.42 | -0.52 |
| Joormann et al. (1) | 2005 | Suppress positive | unspecified | Words | Positive | 4 | 12 | Depression | DEP | 18 | 18 | Retrieval ACC | 0.10 | 0.01 | -0.09 |
| Joormann et al. (2) | 2005 | Suppress negative | unspecified | Words | Negative | 4 | 12 | Depression | DEP | 18 | 18 | Retrieval ACC | 0.53 | 0.30 | -0.26 |
| Küpper et al. | 2014 | | Direct RS | Pictures | Negative | 3 | 10 | Low thought control | Mixed | 12 | 12 | Correct ID | 0.41 | 0.69 | 0.79 |
| Marzi et al. | 2013 | | Direct RS | Pictures | Neutral | 3.5 | 5 | High trait anxiety | ANX | 15 | 15 | Retrieval ACC | 0.49 | 0.86 | 0.33 |
| Nemeth et al. | 2014 | | unspecified | Words | Neutral | 3.5 | 16 | Alcohol abuse | Mixed | 36 | 36 | Retrieval ACC | -0.48 | 0.08 | 0.61 |
| Noreen & Ridout | 2016a | | unspecified | Words | Neutral | 3 | 5 | Dysphoria | DEP | 36 | 36 | Retrieval ACC | 0.26 | 0.38 | 0.20 |
| Noreen & Ridout (1) | 2016b | Aided Suppression | Thought Sub | Words | Mixed | 3 | 8 | Dysphoria | DEP | 18 | 18 | Retrieval ACC | -0.86 | 0.49 | 1.39 |
| Noreen & Ridout (2) | 2016b | Unaided Suppression. | unspecified | Words | Mixed | 3 | 8 | Dysphoria | DEP | 18 | 18 | Retrieval ACC | -0.88 | -0.73 | 0.19 |
| Sacchet et al. | 2017 | | unspecified | Words | Neutral | 3 | 12 | Depression | DEP | 16 | 16 | Retrieval ACC | 0.31 | 0.25 | -0.18 |
| Salamé & Danion | 2007 | | unspecified | Words | Neutral | 3 | 16 | Schizophrenia | Mixed | 23 | 24 | Retrieval ACC | 0.54 | 0.07 | -0.61 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Waldhauser et al. | 2018 | | Direct RS | Pictures | Neutral | 2 | 12 | PTSD | ANX | 11 | 13 | Correct hits | -0.25 | 0.56 | 0.74 |
| Wessel et al. | 2005 | | unspecified | Words | Neutral | 3 | 16 | High dissociation | Mixed | 35 | 33 | Retrieval ACC | 0.27 | 0.20 | -0.05 |
| Zhang et al. | 2016 | | Direct RS | Pictures | Neutral | 3 | 10 | Dysphoria | DEP | 25 | 25 | Retrieval ACC | 0.86 | 0.68 | 0.02 |

**Table 1. Studies included in the meta-analyses.** *Condition* identifies the effect size if the respective study reported more than one. *Instructions* indicate whether the respective study aimed at inducing a specific suppression mechanism. *Instructions* as well as *Material*, *Valence*, *Duration*, and *Repetitions* were included as potential moderators. The *Dependent variable* identifies the respective test used to quantify suppression-induced forgetting. *Clinical sample* identifies the specific clinical or sub-clinical condition investigated in each study, and *Cluster* the broader clinical/sub-clinical sample it was assigned to (e.g., anxiety; see 2.3). *N* indicates the size of the respective samples. *SIF* (= suppression induced forgetting) indicates the within-group effects (standardized mean change with change score standardization, *SMCC*) of the clinical/sub-clinical and healthy control groups as well as for the respective group difference (standardized mean difference, *SMD*). List of acronyms: direct RS = direct retrieval suppression, thought Sub = thought substitution (instructions column); ANX = anxiety, DEP = depression (clinical cluster column); *N* = sample size; DV = dependent variable, correct ID = correct identification, retrieval ACC = retrieval accuracy (DV column).